
Testing, Testing, 1, 2, 3

*Raise More Money
with Direct Mail Tests*

M A L W A R W I C K



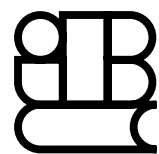
JOSSEY-BASS

A Wiley Imprint

www.josseybass.com

VISIT...

LANZAROTE
Caliente.COM



Other Jossey-Bass publications by Mal Warwick:

How to Write Successful Fundraising Letters, 0-7879-5652-X

Fundraising on the Internet: The ePhilanthropyFoundation.Org's
Guide to Success Online, 0-7879-6045-4

The Five Strategies for Fundraising Success: A Mission-Based Guide
to Achieving Your Goals, 0-7879-4994-9

Ten Steps to Fundraising Success: Choosing the Right Strategy for
Your Organization, 0-7879-5674-0

The Jossey-Bass Nonprofit and Public Management Series

also includes:

Inspired Philanthropy: Your Step-by-Step Guide to Creating a Giving
Plan, 2nd Edition, *Tracy Gary & Melissa Kohner*, 0-7879-6410-7

Relationship Fundraising: A Donor-Based Approach to the Business
of Raising Money, 2nd Edition, *Ken Burnett*, 0-7879-6089-6

Selling Social Change (Without Selling Out): Earned Income
Strategies for Nonprofits, *Andy Robinson*, 0-7879-6216-3

Opening Doors: Pathways to Diverse Donors, *Diana S. Newman*,
Council on Foundations, 0-7879-5884-0

Winning Grants Step by Step, 2nd Edition, *Mim Carlson*, 0-7879-5876-X

Conducting a Successful Capital Campaign, Second Edition,
Kent E. Dove, 0-7879-4989-2

The Grantwriter's Start-Up Kit, *Creed C. Black, Tom Ezell*,
Rhonda Ritchie, Successful Images, Inc., 0-7879-5233-8

Secrets of Successful Grantsmanship, *Susan L. Golden*, 0-7879-0306-X

The Insider's Guide to Grantmaking, *Joel J. Orosz*, 0-7879-5239-9

Hidden Assets, *Diane L. Hodiak, John S. Ryan*, 0-7879-5351-2

Transforming Fundraising: A Practical Guide to Evaluating and
Strengthening Fundraising to Grow with Change,
Judith E. Nichols, 0-7879-4495-5

Fund Raisers: Their Careers, Stories, Concerns, and
Accomplishments, *Margaret A. Duronio, Eugene R. Tempel*,
0-7879-0307-8

The Fundraising Planner: A Working Model for Raising the Dollars
You Need, *Terry & Doug Schaff*, 0-7879-4435-1

Preparing Your Capital Campaign, *Marilyn Bancel*, 0-7879-5247-8

Developing Your Case for Support, *Timothy L. Seiler*, 0-7879-5245-1

Planning Special Events, *James S. Armstrong*, 0-7879-5248-6

Building Your Direct Mail Program, *Gwyneth J. Lister*, 0-7879-5529-9

Planning and Implementing Your Major Gifts Campaign,
Suzanne Irwin-Wells, 0-7879-5708-9

Testing,
Testing,
1, 2, 3

Testing, Testing, 1, 2, 3

*Raise More Money
with Direct Mail Tests*

M A L W A R W I C K



JOSSEY-BASS
A Wiley Imprint
www.josseybass.com

Copyright © 2003 by Mal Warwick

Published by Jossey-Bass

A Wiley Imprint

989 Market Street, San Francisco, CA 94103-1741 www.josseybass.com

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400, fax 978-750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, 201-748-6011, fax 201-748-6008, e-mail: permcoordinator@wiley.com.

Jossey-Bass books and products are available through most bookstores. To contact Jossey-Bass directly call our Customer Care Department within the U.S. at 800-956-7739, outside the U.S. at 317-572-399386 or fax 317-572-4002.

Jossey-Bass also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

This work was previously published in 1992 by Strathmoor Press as *You Don't Always Get What You Ask For: Using Direct Mail Tests to Raise More Money for Your Organization*.

Library of Congress Cataloging-in-Publication Data

Warwick, Mal.

Testing, Testing, 1, 2, 3: raise more money with direct mail tests/Mal Warwick.

p. cm.—(The Jossey-Bass nonprofit and public management series)

Includes index.

ISBN 0-7879-6712-2 (alk. paper)

1. Direct-mail fund raising. 2. Testing. I. Title.

II. Series.

HV41.2.W379 2003

2003001628

Printed in the United States of America

FIRST EDITION

PB Printing 10 9 8 7 6 5 4 3 2 1

Contents

Perspectives and Tables	xi
Preface	xv
The Author	xxi

PART I

The Rules of the Road	1
1 Testing, Testing, Testing—Until You’re Blue in the Face	3
2 The Initial Test Mailing	9
3 What’s Worth Testing (and What Isn’t)	15
4 No Size Fits All	21
5 What You See—Versus What You Get	25
6 Tests or Rollouts?	29
7 Results Worth the Wait	35
8 Reading Your Results	39
9 Understanding Confidence Levels	45
<i>Dan Suzio</i>	

10	The Value of Guesswork	55
11	Testing Tips from Major Mailers	59
12	Testing Is Fundamental	65
13	Comparing Costs	71

PART II

Real-World Results 77

14	Signatures	79
15	Teasers	83
16	Brochures	87
17	Window Envelopes	91
18	Mailing for the Environment (I)	95
19	Mailing for the Environment (II)	99

PART III

Testing in Action 103

20	Inserts	107
21	Postage	117
22	Outer Envelopes	123
23	Outer Envelope Teasers	133
24	BREs	143
25	Letter Length	149
26	Gift Amounts	155
27	Gift Strings	181
28	Credit Card Options	185

29	Miscellaneous Tests	189
30	Spending Money to Raise Money	223
	Conclusion: Seven Guidelines for Successful Testing	227
	Glossary	235
	Recommended Reading	237
	Index	241

Perspectives and Tables

Perspectives

Shibboleths?	7
Package Cost?	12
How Much?	17
Out with the Old?	24
Donor Involvement?	28
Pop Quiz	32
Write Long or Short?	38
What Counts?	43
Significant?	49
Once Again, How Much?	57
Response Devices?	62
Credit Cards?	68
Courier Type?	74

Tables

4.1	Results of Three Similar Donor Upgrade Tests	23
5.1	Testing the Validity of <i>Testing</i>	27
7.1	Eight Months of Revenue Receipts in a Donor Acquisition Test	37
9.1	Hypothetical Initial Test Results	46
14.1	Testing the Impact of a New Letter-Signer	80
15.1	Testing Teaser Versus No-Teaser Carriers for Two Organizations	85
16.1	Testing Brochure Versus No-Brochure Packages for Two Public Interest Organizations	89
17.1	Testing Window Versus Closed-Face Envelopes	92
18.1	Testing 4-Sheet Versus 2-Sheet Letters with Identical Text	97
19.1	Testing the Value of an Environmental Message on Environmental Donors	101
20.1	Testing Inserts	109
21.1	Testing Postage	119
22.1	Testing Outer Envelopes	125
23.1	Testing Outer Envelope Teasers	135
24.1	Testing BREs	145
25.1	Testing Letter Length	151
26.1	Testing Gift Amounts	157

27.1	Testing Gift Strings	182
28.1	Testing Credit Card Options	186
29.1	Testing Other Package Elements: Logos, Surveys, Reply Devices, Typefaces, Booklets, and More	191
30.1	Testing Live Stamp Versus No Stamp Reply, Acquisition Mailing, at 16 Weeks	224
30.2	Testing Live Stamp Versus No Stamp Reply, Resolicitation Mailing, at 40 Weeks	224

*This book is dedicated to Nancy—
Who brings joy to my life
And to everyone she meets.*

Preface

Time and time again, experience shows that subtle changes in the copy, design, or target audience of a fundraising mailing can dramatically affect that mailing's outcome.

Ask any direct mail or telephone fundraiser. You'll probably hear stories of response rates or average contributions that doubled—or dropped by half—when the art on the outer envelope was slightly changed, or a new list was used, or an inducement was offered to those who sent in gifts quickly.

This is the stuff of which breakthroughs in fundraising are made—not the incremental, mailing-by-mailing improvements but those quantum leaps that open up new vistas for the fundraiser.

But breakthroughs don't come easily or often. Like qualitative changes in other fields, huge leaps in fundraising performance come rarely. Moreover, they come about only when you look for them. Without an organizational commitment to continual testing—exploring new lists, new offers, and new packages as well as significant variations on the existing lists and packages—breakthroughs in fundraising are unlikely to come.

In any case, changes such as these, whether subtle or dramatic, are direct mail and telephone fundraisers' bread and butter. Certainly, whether we are in-house fundraisers or external consultants, we earn our keep only when we bring knowledge to bear as we engineer changes to improve results.

But how can we *know* whether the changes we make will be for the better? The way we prove our case—the *only* way to know—is by testing.

Direct response testing is far easier to define than to execute. Most of us take the need to test for granted. But few of us test consistently, or analyze test results thoroughly, or always follow through by acting on the lessons we've learned.

This book is one fundraiser's attempt to cast light on some of the most important (if often obscure) realities of testing: its true value (not just what we might imagine it to be), its character in practice (not theory), and its limitations (which are many).

Who Should Read This Book

This book will be of greatest immediate value to those who are directly engaged in direct mail fundraising. You'll learn why, how, and when to test, and what you can expect from testing. But I believe that *Testing, Testing, 1, 2, 3* will also be of great value to others:

- To the vice president or director of development who needs to understand the potential and the limits of direct mail.
- To other development staff members who wish to gain an appreciation of how direct mail fundraising really works.

- To trustees or board members who want to learn what's possible, and what isn't, in direct mail fundraising, and understand better the complex demands of professional fundraising.
- To direct marketers who wish to gain perspective on how direct mail fundraising is different from their craft.

Contents of This Book

Testing, Testing, 1, 2, 3 is organized into three parts:

Part One, “The Rules of the Road,” lays out in thirteen chapters the advantages—and the drawbacks—of testing and relates the fundamental rules of the process. *Perspective* sections in these chapters offer examples of actual tests that illuminate these rules.

Part Two, “Real-World Results,” analyzes the results of tests probing the usefulness of five direct mail devices and issues that fundraisers often use.

Part Three, “Testing in Action,” presents the results of 286 separate direct mail fundraising tests, organized by ten key topics.

In the final chapter, I suggest points of departure for fundraisers eager to strike out on their own.

The material in this book is drawn largely from three sources:

- A little book (now long out of print) that I wrote for Strathmoor Press in 1992, titled *You Don't Always Get What You Ask For: How Direct Mail Tests Can Help You Raise More Money for Your Organization*. Along with other material, that book contained (sometimes edited and expanded) nineteen of the monthly columns I wrote under the title “The Warwick File” for *The NonProfit Times*. All nineteen come from the first two of my seven years of producing that column.

- Testing data that have appeared in my bimonthly print newsletter over the years. This newsletter, which I've been writing and editing under one name or another since 1986, now sports the immodest title *Mal Warwick's Newsletter: Successful Direct Mail, Telephone & Online Fundraising*.
- The results of 286 individual direct mail tests my colleagues and I have conducted in recent years for thirty-one of our clients. I've adapted these results from what I and others in my organization laughingly call our agency's "Testing Bible."

Altogether, this material constitutes, to my knowledge, the *only* large-scale compilation of direct mail fundraising test results committed to print for almost two decades.

Now, I don't pretend this book is a comprehensive treatment of the subject of testing new approaches in direct mail fundraising. Some important issues are dealt with only in passing, if at all. For example, this book presents relatively little material on the subject of list testing—and as you'll soon learn (if you don't already know from long experience), the audience is the *most* important thing to test.

In other books I have dealt at length with two closely related issues that also must be confronted before any decisions are made about testing: *acquisition cost* and the *lifetime value of a donor*. After all, it makes little sense to monkey around with low-end tests designed to tweak the response rate when your organization really needs higher average initial gifts. So if these two phrases cause you to scratch your head in puzzlement, you would be well advised to read up on them before putting the lessons in this book into practice.

At the present time, fundraisers face a near-vacuum when they try to find current, reliable information on the realities of

direct mail testing. This volume is my entry into that field—a modest effort to provide direct mail and telephone fundraisers with practical insights into a subject that's of central importance to their craft. I hope it helps you.

Acknowledgments

Most of the test results cited here—including all those noted in Part Three—were drawn from the experience of my direct mail fundraising firm, Mal Warwick & Associates, Inc.

I am especially indebted to three people at Mal Warwick & Associates, all of whom have been instrumental in the preparation of this book:

Dan Suzio, whose nominal title at Mal Warwick & Associates is network administrator but who plays a much more central role in our work for our clients because he manages all our statistical reporting systems. Whenever I have a question about testing methodology or the interpretation of test results, I talk to Dan. He has been our teacher over the years, instructing us in the finer points of the process of testing, and he contributed Chapter Nine, “Understanding Confidence Levels,” to this book.

Dan Weeks, who compiles and updates our “Testing Bible,” much of which I’ve incorporated into Part Three. Dan’s diligence and continuous attention to detail have made it possible for me to relate so many real-world examples in this book.

Stephen Hitchcock, president of Mal Warwick & Associates, for his unfailingly insightful and cheerful assistance in collecting and interpreting the test data I refer to in these pages and for contributing the article that I’ve adapted into Chapter Two, “The Initial Test Mailing.”

I'm also indebted to Tom Collins, who wrote the long, thoughtful letter from which I quote extensively in Chapter Eight.

Despite all the help and advice I've received, the point of view and the conclusions drawn are mine alone. And if you disagree with them, I invite you to drop me a line at mal@malwarwick.com. But I give you fair warning now: to *prove* that I'm wrong, you'll have to do what every direct mail fundraiser should do anyway, as a matter of course—test it, and prove it for yourself!

*Berkeley, California
January 2003*

Mal Warwick

The Author

MAL WARWICK is a consultant, author, and public speaker who has been involved in the not-for-profit sector for nearly forty years. He is the founder and chairman of Mal Warwick & Associates, Inc. (Berkeley, California), a fundraising and marketing agency that has served nonprofit organizations since 1979, and of its sister company, Response Management Technologies, Inc., a data processing firm for nonprofit organizations. He is also a cofounder (with Nick Allen) of Donordigital.com LLC (San Francisco), which assists nonprofit organizations on-line, and a cofounder of Share Group, Inc. (Somerville, Massachusetts), the nation's leading telephone fundraising firm.

Mal has written or edited fourteen previous books of interest to nonprofit managers, including *The Five Strategies for Fundraising Success*, *Fundraising on the Internet*, and the standard texts *Raising Money by Mail* and *How to Write Successful Fundraising Letters*. He is the editor of *Mal Warwick's Newsletter: Successful Direct Mail, Telephone & Online Fundraising* and is a popular speaker and workshop leader throughout the world.

Among the hundreds of nonprofits Mal and his colleagues have served over the years are many of the nation's largest and most distinguished charities as well as six Democratic presidential

candidates and scores of small, local, and regional organizations. Collectively, Mal and his colleagues have been responsible for raising at least half a billion dollars—largely in the form of small gifts from individuals.

Mal is an active member of the Association of Fundraising Professionals and the Direct Marketing Association Nonprofit Federation, and he served for ten years on the board of the Association of Direct Response Fundraising Counsel, two of those years as president. He also serves on the international board of the Resource Alliance, which organizes the annual International Fundraising Congress, and is its U.S. country representative.

Mal was a Peace Corps volunteer in Ecuador for more than three years. He has been active in civic affairs in his home community of Berkeley since the 1960s. Early in the 1990s, he cofounded the Community Bank of the Bay, the nation's fifth community development bank, and the Berkeley Community Fund, where he remains active on the board. He is a cofounder of Business for Social Responsibility and served on its board during its inaugural year. In 2001, he was elected to the board of the Social Venture Network and now serves as chair.

He is the grandfather of Dayna, Iain, Matthew, Gwen, Andrew, Kaili, and Benjamin.

Part I

The Rules of the Road

The following thirteen chapters lay out the basic principles of direct mail testing. You'll learn what a test is, and what it isn't. And I'll give you some pointers about why to test, how to test, and what to test.

Testing, Testing, Testing— Until You're Blue in the Face

Every once in a rare while, you'll hit the nail on the head with a powerful marketing concept—a seemingly perfect marriage of direct mail package and lists—at the very beginning of a new program. Your *control* package—the standard-bearing donor acquisition appeal you can keep mailing—may emerge more or less wholly formed from a successful pilot mailing.

I've seen that happen more than once.

But much more often it takes two or three tries, and months or years of step-by-step refinements, to produce a workhorse prospecting package that will build a large, responsive donor file for your organization. And then, sometimes quite suddenly, that package will “die,” and you'll have to come up with another one very quickly.

A donor acquisition program might not hit its stride until its third or fourth year or even later. The length of this process may

relate to the organization's gradual accumulation of credibility and public recognition, to some shift in public sentiment about the issues the organization is addressing, or to changed attitudes in the organization's top management. But chances are it has a lot more to do with the cumulative value of thoughtful, systematic testing.

Now, here's what I mean by a *test*:

You might, for example, want to know whether the response to your prospect package will rise significantly if you suggest a lower minimum gift in the letter and on the response device. Currently, the package urges a contribution of \$25 or more; you decide to test whether suggesting a \$15 minimum will lift the response rate sufficiently to lower the overall donor acquisition cost (and without forcing the average gift to drop in proportion).

To determine this—reducing the question to its simplest form—you would take two equal and statistically equivalent groups of names from the same pool of prospect lists, mailing one group a package with the \$25 suggested minimum *ask* and the other with the \$15 ask.

The theory behind direct mail testing such as this is that applying scientific principles will enable you to discover the perfect combination of ask, offer, package, and postage—and ride off into the sunset, forever financially secure.

The reality's a little different.

For one thing the possible tests in a direct mail fundraising program are innumerable. It's sometimes excruciatingly difficult to figure out what's important to test and what isn't—particularly when your testing budget is strictly limited. In a program of modest scope it's especially important to test only those elements that are *significant*. Too much testing can eat up all the profits from your direct mail program.

A wholly new acquisition package built around a new marketing concept is likely to be a significant test. Most of the time, so are meaningful variations in the benefits offered to donors. But even such variations might be insignificant under some circumstances.

Another problem with the proposition that rigorous testing will show the way to optimal results is that it's often difficult to design tests that are statistically meaningful. In most situations it requires many tens of thousands of names to establish confidence in the results of such things as price or postage tests. And even when you're able to test in these relatively large quantities, you'd be well advised to view the results with caution—no matter what the statisticians tell you.

Time and again I've seen results vary 20 percent or more between even very large samples when the variation was supposed to be no greater than 5 or 10 percent. According to experts in statistics, variations this great shouldn't occur more often than 1 to 5 percent of the time. I suspect they *often* occur, and so do some other direct mail fundraisers.

To test the validity of this proposition, on several occasions my colleagues and I arranged what we were assured were rigorous conditions: we split large groups of letters into two statistically identical groups to see what effect random variation might have. There was *no* difference between the two groups in package, postage, list, or any other controllable factor. All we were testing was the extent to which two equal prospect samples would produce equal results when simultaneously mailed identical packages. On both occasions, random variation had a greater effect than statistical theory said it should. Other consultants have had similar experiences.

No doubt there's some obscure theory of statistics that can explain away even these anomalous experiences. Rather than chase

it down, however, I prefer to hold my clients' losses to a minimum and trust as little as possible in statistics—except when mailing quantities and testing budgets are big enough to add an extra margin of safety.

Why, then, when individual results may not be statistically reliable, should we bother to test such things as ask amounts, teasers, or postage? The answer is simple: by and large, *testing works*.

Testing works when we fundraisers test with *expectations* and then incorporate the results of our tests into future packages when our expectations are confirmed. Over time—sometimes after retesting vital elements two or three times—we're likely to learn valuable lessons that will enable us to upgrade a tolerably workable control package into one that is truly responsive.

Through testing—lifting response by 10 percent with this little change and 5 percent with that one—we may eventually cut the donor acquisition cost by 50 or 75 percent. Testing may, and often does, bring new life into a prospecting program by permitting us to expand the mailing volume and build large donor files relatively quickly. In large donor resolicitation programs, testing can also sometimes dramatically increase net revenue by establishing the most cost-effective use of postage, packaging techniques, suggested contributions, and the like. In this approach to direct mail prospecting, everything has the potential to be tested. As the following perspective section for this chapter suggests, even what everyone accepts as true may benefit from a reality check.

Testing is not limited to large direct mail programs. Almost any mailing of 10,000 names or more affords an opportunity to test. Smaller quantities limit the options and make testing a longer process. They demand extra care and greater patience, but the payoff can be just as dramatic in the end.

PERSPECTIVE: Shibboleths?

Conventional wisdom in direct mail fundraising maintains that a credit card option on the response device will depress results. It also says that a real stamp (a *live stamp*) on the package will boost response in comparison to the response to a package with a printed postal indicia or even a postage meter imprint.

'Tain't necessarily so. Not anymore, anyway. In a recent test for a client, my firm ran a head-to-head test of a check or credit card offer versus a check-only offer. There was no significant difference in response. Clearly, this is consistent with the experience of many major nonprofit mailers, who now routinely accept gifts by credit card. The same could be true for your organization—maybe.

In a test for another client, we compared a live, nonprofit bulk stamp with a postage meter imprint, mailing 30,000 packages in each fashion and expecting the stamp to outpull the meter, at least by a little. But the opposite proved to be the case. The meter out-pulled the live stamp by 170 to 138 returns—a difference that was statistically significant. In other words these findings suggest it's not worth paying the extra cost of affixing a live stamp—at least it wasn't for that client at that time.

So, what conclusions can fundraisers draw from knowing that test results may not support what they have been assured is typical? Three conclusions, I believe:

1. Don't take conventional wisdom for granted.
2. Things may change over time.
3. Continual testing pays.

And remember, a *test* is just as real a mailing as one that's called something else. One nonprofit's small test is another's huge prospect mailing. But both have to be evaluated in terms of costs and benefits.

Be sure, however, that you don't fall prey to the common but costly mistake of evaluating test results only in terms of their actual costs. Testing, as you will see in some of my examples, is often expensive. It may involve added creative, production, and even management costs—all of which need to be factored out when analyzing the outcome. Otherwise, you'll be comparing expensive apples to cheap oranges. Testing costs should be considered part of the overhead of your nonprofit's direct mail fundraising program as a whole. In fact, it's wise to add a line allocated specifically to testing to your direct mail budget.

Now, once you've designed and conducted your test, you've got to read and use the results, and that may be more easily said than done. The results from your test packages may conflict with the findings of earlier tests or with each other. The results will also be open to varying interpretations, depending on each interpreter's goals for the test. Furthermore, errors made in the *lettershop* (where your mailing is addressed and assembled) may call the validity of the results into question.

But the *biggest* danger in testing is that the uncertainty caused by equivocal results—or the anxiety raised by poor results—might tempt you to delay future mailings.

Important as your test results might be, the chances are that if you are trying acquire more donors for a productive, ongoing direct mail fundraising program, you'll be better off mailing blind than waiting for a definitive reading from tests—because your biggest mistake may be not to mail at all.

The Initial Test Mailing

To start off on the right foot in direct mail testing, follow these nine guidelines for your *first* mailing:

1. The most important factor to test at the outset is whether or not direct mail acquisition of donors is even feasible for your organization. Direct mail fundraising won't work for every organization—but you've probably got to try to find out for sure.
2. To test the feasibility of direct mail for your organization, it's likely you'll need to mail at least 30,000 letters—and preferably 50,000. In some circumstances, smaller mailings may make sense. But if your budget allows, you'll find an even *larger* mailing may be advantageous because

- it allows you to collect more information and do it more quickly—potentially saving a lot of time and money later.
3. In your initial mailing, you should test various types of prospect lists to see which ones will work best for you. You'll also get a surer reading of your future potential to mail in large enough quantities to sustain an ongoing acquisition program.
 4. Generally, the lists tested in your first mailing should be the ones you feel are most likely to do well—lists of people who already donate to organizations like yours. But you should also test lists with larger donor universes and lists of nondonors, such as magazine subscribers. This way, if *any* of these less-promising lists do prove fruitful, you'll have lists with *rollout* potential for future acquisition mailings.
 5. Usually, 5,000 is the minimum number of names you need to randomly select from a single list in order to make a valid test of that list. Thus a 50,000-piece mailing will let you test ten different lists. (If the rate of response to your 5,000-piece mailing is 1 percent, a typical list will yield 50 returns; statistical purists, however, often insist on a minimum of 100 or even 200 returns to ensure valid results. That's why I also suggest that you try to include one or two lists of, say, 7,000 to 10,000 names each—and keep your fingers crossed about the others!)
 6. It's almost always dumb not to merge-purge your lists in order to reduce the number of duplicate names and eliminate names of your current donors. A merge-purge won't just reduce your costs and complaints from donors. It will also help ensure that results aren't "polluted" by the presence of these duplicates and existing donors.

7. Restrict your first mailing to just *one* package, sent to all the lists. This will keep your design, printing, and postage costs to a minimum.
8. If you have enough money for an initial test mailing of 60,000 pieces or more, you can test two versions of the test package. Use this opportunity to test the ask.
9. To maximize your chances of success in your initial test, I urge you to spare no expense on this first piece. Then in future mailings you can test to see what corners can be cut. As the legendary direct mail copywriter Bill Jayme said in a 1977 *Advertising Age* interview: “I believe in going for broke. If I were launching a new venture, I’d pack everything into the initial package. . . . And then, if the package was successful, I’d test down . . . cutting back on costs until returns begin to hurt. If you start with an economy package and it bombs, you’ll always wonder what would have happened if you had gone for broke in the first place.” (And as the perspective for this chapter suggests, after you’ve tested down, even the resulting successful penny-pinching package should be periodically retested.)

If the results of your initial test mailing are encouraging, you’ll be ready to launch a second mailing.

At this point the rules of testing start to change.

In your *second* acquisition mailing, if you haven’t already done so, it’s important to test the ask. In fact, the following question is important enough that you may be well advised to test it at every opportunity: What initial contribution will generate both the most income (requiring a high average gift) *and* the most new donors (because the percentage of respondents is high)? One large, national environmental organization has had success asking for only \$15,

PERSPECTIVE: Package Cost?

When does it pay to spend more money on a package rather than less? This question can cause considerable consternation among direct mail fundraisers—especially those of us with fanatically cost-cutting clients or bosses. It often doesn't pay to add components or improve the production quality of a direct mail package.

But there are exceptions to this generalization—lots of them, in truth. Here's an example. For some time, one of my firm's clients had been successful mailing its control package, which was stripped down to essentials. Then a colleague decided to test adding a costly sticker. This increased the package cost from approximately \$0.36 apiece to \$0.41—a substantial boost. The test compared results from two statistically equivalent panels of about 40,000 prospects each.

The table for this example (like many of the tables to come) shows, for each type of package (Pkg.), the total number of gifts received (Total Gifts), the percentage of people who responded (% Resp.), the average gift (Avg. Gift), the revenue per thousand letters mailed (Rev./1,000), and the cost per thousand letters mailed (Cost/1,000). Note that in most of these tables the average

whereas a smaller, regional organization asks for a first-time gift of \$50. Some groups prefer to obtain more donors and don't mind a lower average donation. Others would rather maintain files of fewer donors among whom many are larger contributors.

Keep in mind, however, that if you're building a direct mail program from scratch, you're better off acquiring as many names as possible. This will give you the opportunity to trade more names with other like-minded organizations and to solicit these

gift takes into account only gifts of less than \$100, so that a few large, outlying gifts will not skew the average upward (as I explain in more detail later in the perspective section for Chapter Eight). However, the number of total gifts does include gifts over \$100, so as to accurately reflect the total number of new donors.

Sticker Test Results.

Pkg.	Total Gifts	% Resp.	Avg. Gift*	Rev./ 1,000*	Cost/ 1,000
Sticker	393	0.98%	\$30.38	\$275	\$412
No sticker	297	0.74	29.89	200	355
Difference with sticker	+96	+0.24	+0.49	+75	+57
% difference with sticker	+32%	+32%	+2%	+38%	+16%

* These figures take into account only gifts of less than \$100. There was substantial revenue from \$100+ gifts in addition.

I don't know about you, but anytime I can increase revenue by 38 percent by spending only an additional 16 percent, I'll jump at the chance!

additional donors yourself in the future. So at least at first you'll probably want to opt for more donors by asking for a smaller first-time gift, even if that results in a lower average gift.

Other important things to test in the early days of a new direct mail program are the following:

- Involvement devices such as petitions, postcards, quizzes, and surveys
- Size of the outer envelope

- Front-end premiums (those sent with the package) like stickers, stamps, or decals
- Back-end premiums (those sent after a gift is received) such as calendars, notecards, framed pictures, or hip pouches

Remember, the value of any test depends on strict adherence to testing standards:

- Randomly select your *test panels* (that is, the names included in each distinctly coded group in the test).
- Test only one thing at a time.
- Mail the package (or packages if you are testing two versions) to all test panels simultaneously and from the same location.
- Make sure the copy in your acquisition package stays away from short-term issues and programs. Your package needs to stay pretty much the same for a time so you can use it as your control in at least one subsequent test.

And here's a final thought. If your initial test suggests that direct mail acquisition works for your organization, you've only gotten started. You should test diligently *every* time you mail!

What's Worth Testing (and What Isn't)

In direct mail fundraising, some of the things we commonly test often make a big difference. Others rarely do.

The *difference* I'm talking about is defined in terms of *cost effectiveness*—taking into account both higher or lower *costs* and a higher response or average gift (higher *income*). After all, the purpose of testing is to find variations in the list selection or the package that will improve your net results in later mailings, to increase the cost effectiveness of your direct mail or telephone fundraising program over time. It's important to keep your eye on *both* costs and income.

In considering the guidelines that follow, please remember, in the immortal words of my fellow Berkeleyan Wavy Gravy, “for every exception, there's a rule.” You would be well advised *not* to take my word for any of this. *Testing* is the name of this game.

Which is another way of saying that you've gotta find out for yourself what works for your fundraising program.

High-Potential Elements

Some elements in a direct mail fundraising campaign are most likely to make a difference on the order of 20 percent—or even more—in your net returns. In other words, if anything's going to give your mailing program a big lift, it's probably one or the other of the following ten items.

- Testing *prospect lists*!
- Suggesting a higher or lower minimum gift (see the sample test in the perspective)
- Conducting telephone follow-up
- Offering a product (premium) to those who give
- Enclosing a free product (premium) in the appeal
- Using an involvement device, such as a survey or petition
- Designing a new package
- Redesigning the outer envelope
- Making a sweepstakes offer

But be very cautious if you elect to acquire donors by using a sweepstakes offer or offering a product in exchange for an initial gift. Your response rate may indeed be higher. But respondents to sweepstakes or premium offers may lack interest in your cause or organization, and their value over the long haul may be substantially less than that of donors or members acquired without such inducements to greed. You may also encounter difficulties with the IRS or the U.S. Postal Service, neither of which is a big fan of non-profit mailings that smack of commercialism.

Moderate-Potential Elements

If they affect your results at all, the following ten items are likely to improve—or decrease—your net returns by no more than 10 or 20 percent.

- Changing the type of postage
- Using a celebrity endorsement
- Altering letter copy
- Personalizing the package

PERSPECTIVE: How Much?

Direct mail testing can be frustrating. All too often, the results vary little no matter what subtle changes you make in a package. A long run of meaningless tests may be almost enough to make you wonder if testing is really worthwhile after all.

Then something turns up in a routine test that restores your faith in the process.

More often than not, I've found, a test that actually produces statistically significant results is likely to involve a change in the minimum suggested gift amount.

For example, in one recent test for a client, my colleagues chose to test the value of the \$35 initial gift ask that was incorporated into the control package. They mailed three different versions to three test panels of 30,000 prospects each, testing the \$35 control against two alternative formats for a \$26.75 minimum ask. In one alternative version the suggested amount was highlighted with a hand-drawn circle on the response device; in the other the same amount was suggested but not highlighted. The table illustrates what happened.

Alternative Ask Test Results.

Pkg.	Total Gifts	% Resp.	Avg. Gift*	Gross Rev.*
\$35 control	143	0.48%	\$26.61	\$3,692
\$26.75 circled	171	0.57	25.20	4,007
\$26.75 not circled	201	0.67	25.31	4,607
Difference between \$26.75 not circled and control	+58	+0.19	−1.29	+915
% Difference between \$26.75 not circled and control	+41%	+40%	−5%	+25%

* These figures take into account only gifts of less than \$100.

The difference in response between the \$35 control and the circled \$26.75 request—just 28 gifts out of a total of 314—is not statistically reliable. But the results for the \$26.75 ask *without* a circle for emphasis are entirely convincing: there is every likelihood that this version will generate significantly more revenue than the control by substituting the \$26.75 ask for the \$35 ask in the control package.

Such wide variation in response is the exception rather than the rule in testing. But when it occurs, it's important to pay attention!

- Mailing to special list selects (for example, to women only or to prospects located only in certain zip codes)
- Mailing to a different geographical region
- Reformatting the package
- Applying a live stamp to the reply envelope
- Highlighting the ask on the reply device
- Mailing in certain months but not in others (*seasonality*)

Low-Potential Elements

More often than not, the following six long-time favorite candidates for testing will make a net difference in direct mail results of less than 10 percent. In other words, I suggest you think twice before executing tests of these items—unless you mail in substantial quantities and have a large testing budget.

- Altering ink color
- Altering letter or envelope stock
- Changing from a business reply envelope to a “Place Stamp Here” reply envelope or vice versa
- Changing the letter signer
- Making the letter longer or shorter
- Changing the postage from stamp to indicia, or indicia to meter, and so on

If you doubt these guidelines—or if some good reason leads you to test one of the propositions I've said is unlikely to make much of a difference—please keep the following in mind: the *smaller* the difference you expect to see in the results, the *larger* the test panel you'll have to select. Otherwise, the difference in the numbers of returns between the test and control panels isn't likely to be large enough to be statistically significant.

No Size Fits All

With postal and production costs constantly on the rise, donor upgrading is an increasingly vital key to successful direct mail fundraising for many nonprofits. Upgrading needn't wait until a donor is on board. After all, if you can recruit her with a higher initial gift, half of the upgrading battle may already be won. That's the logic of a device that's often used in the more venturesome direct mail donor acquisition programs: highlighting a higher-than-minimum suggested initial gift on the response device.

In theory, drawing attention to a higher gift level causes the average contribution to rise. The countervailing risk is that fewer donors will respond, as some may be put off or be unable to meet the higher "minimum."

My direct mail fundraising agency conducted almost identical tests of this proposition for three of its clients. In the control packages mailed in all three acquisition programs, the first gift level suggested was \$25 and the next level was \$35. On a test version of each of the three reply devices, the consultant circled the \$35 box and wrote by hand below the string of suggested gifts, “Your \$35 contribution will really make a difference.” This wording, and the circle around the \$35 box, were the only differences between the test and control packages in this research project.

The mailings conducted for each of these three organizations sent packages to 50,000 statistically identical names and addresses; 25,000 received the test version and 25,000 the control version.

In the mailing for Organization A, the \$35 upgrade improved overall results only marginally (and thus not enough to trust). The average gift on the upgrade version of the package was \$22 as opposed to \$21 and the rate of response dropped slightly (see Table 4.1). The net effect was a lift of \$7 per thousand letters mailed—not statistically significant.

In the other two mailings, however, the upgrade test had the net effect of *downgrading* results.

For Organization B, the control package outpulled the test version by \$49 per thousand letters mailed, or 18 percent—a solid and significant difference. Although the upgrading device lifted the average gift, it caused response to fall.

For Organization C, the control package bettered the test version by \$33 per thousand packages mailed, or 13 percent (apparently a lot, but probably not enough for the organization to be completely confident that it would happen a second time). Here, too, the average contribution rose, but the rate of response was markedly lower.

TABLE 4.1. Results of Three Similar Donor Upgrade Tests

Pkg.	Total Gifts	% Resp.	Avg. Gift*	Rev./ 1,000*
\$35 test, org. A	325	1.30%	\$22	\$298
\$25 control, org. A	340	1.36	21	291
% difference with test	−4.6%	−4.6%	+5%	+2%
\$35 test, org. B	262	1.05	25	266
\$25 control, org. B	332	1.33	23	315
% difference with test	−27%	−27%	+9%	−18%
\$35 test, org. C	177	0.71	35	247
\$25 control, org. C	245	0.98	28	280
% difference with test	−38%	−38%	+25%	−13%

* These figures take into account only gifts of less than \$100.

The most obvious conclusion from these results is that efforts to upgrade direct mail donors' gifts don't necessarily work. In none of these three attempts to elicit higher initial contributions was there any improvement in overall results. It's true enough that in all three cases there was an increase in the average contribution—but the response rate also fell every time. The net effect was at best a wash.

Moreover, these examples are a sobering reminder of an even broader lesson for all of us engaged in direct mail fundraising: testing is still the name of our game. The results of these virtually identical tests for three different organizations yielded significantly different results. What works for *one* nonprofit may not be valid for another. Even more sobering is the following perspective, which reveals that what works best for an organization at one time may not work best for the same organization at another time.

PERSPECTIVE: Out with the Old?

For years one of my agency’s clients had flourished using a dramatic 6-by-9-inch outer envelope (OE) in its donor acquisition control package. Tests of all sorts of alterations and embellishments almost always failed to improve its performance.

Then a colleague who manages the program elected to test a #10 OE against the 6-by-9-inch control. The table displays what happened in a head-to-head test of two 50,000-name segments.

Outer Envelope Test Results.

Pkg.	Total Gifts	% Resp.	Avg. Gift*	Rev./ 1,000*	Cost/ 1,000
#10 OE test	466	0.93%	\$19.63	\$173	\$366
6" x 9" OE control	371	0.74	17.85	124	394
Difference with #10 OE	+95	+0.19	+1.78	+49	−28
% difference with #10 OE	+26%	+26%	+10%	+40%	−8%

* These figures take into account only gifts of less than \$100.

Who woulda thunk it? Is it possible that prospective donors had become immunized to the control package through repeated exposure and that the downsized carrier was enough of a change to recapture their attention? Who knows?

The ways of the market are mysterious indeed!

In any case this represents one of those rare examples of an unequivocally successful direct mail test—proving that sometimes you really can have your cake and eat it too. You can’t sneeze at a 40 percent lift in revenue—with an 8 percent drop in cost!

What You See— Versus What You Get

For direct mail fundraisers, testing is the stuff of which great success stories are built, brick by brick. Uncounted millions of dollars have been raised as a result of careful, step-by-step improvements in direct mail list selection, suggested gift levels, premium offers, packaging, postage choices, and myriad other elements.

But testing has its limits. And sometimes those limits can be downright discouraging.

One recent test conducted by a fast-growing West Coast environmental organization illustrates just how confusing and seemingly contradictory direct mail test results can be. In a 149,000-piece membership acquisition mailing, the organization chose to test the impact of using a newly redesigned logotype. It selected at random two panels of 30,000 addresses each. One panel (the control) was

mailed an acquisition package using the old logo; the package sent to the second panel (the test) used the new logo.

For an outsider there was nothing very interesting in that. What made the test noteworthy was this: following the advice of its direct mail fundraising counsel, the organization selected a *third* panel as a double-check. This third panel was a *null test*. It also consisted of 30,000 addresses, and it was statistically indistinguishable from the control panel. All 60,000 addresses in the control and null panels were sent *identical* letters employing the organization's old logo. The packages to all three panels were mailed simultaneously.

The null test panel outpulled the control panel by a larger margin than the difference between the control and the test! The 30,000 letters sent to the null test panel generated 319 gifts, more than the control, which yielded just 281—even though the packages mailed to these two groups were totally identical.

But not so fast. As a statistician would tell you, in this test the difference between 281 and 319 is not very meaningful. The difference between the new logo panel and the old logo null panel (272 versus 319 gifts)—is more significant. But even that difference doesn't ring the bell of statistical certainty. In other words, from a statistical perspective, none of the results of this test (summarized in Table 5.1) were meaningful.

But, heeding instinct and common sense as well as the precepts of statistics, you are nevertheless entitled to wonder what produced such a seemingly big difference between two statistically identical groups sent the same (old logo) package. This clearly seems an anomaly. What, then, might explain it?

TABLE 5.1. Testing the Validity of Testing

Pkg.	Total Gifts	% Resp.	Avg. Gift*	Gross Rev.*
Old logo, control panel	281	0.94%	\$25.05	\$5,939
New logo, test panel	272	0.91	22.39	5,490
Old logo, null panel	319	1.06	25.42	6,808
% difference between null and control	+14%	+13%	+2%	+15%
% difference between null and test	+17%	+16%	+14%	+24%
% difference between control and test	+3%	+3%	+12%	+8%

* These figures take into account only gifts of less than \$100.

Take your pick of the following three explanations:

- The lettershop or the postal service messed up—dropping the test panels on different days, perhaps, or improperly packaging or processing them.
- Somebody counted wrong.
- Statistics is not the exact science it's cracked up to be. What's supposed to happen 95 out of 100 times actually happens less often than that.

Whatever the reason, these unpredictable results should drive home a fundamental lesson of direct mail testing: if it's important enough to test once, test it *again* to be sure. (In the example, a retest might be in order because the old logo outperformed the new one in both head-to-head measurements.) Because what you see is *not* necessarily what you'll get. The perspective for this chapter presents another instance of this truth, in this case seeing a difference where none was expected.

PERSPECTIVE: Donor Involvement?

As a loyal fan of involvement devices, I often test these beguiling little contrivances in hopes of encouraging closer relationships with donors as well as generating greater response from mailings.

Over the years, I’ve tried just about every legal and ethical device imaginable. But like so many others in the business, I’ve come back again and again to two time-tested options: petitions and surveys. How interesting it was, then, when one of my colleagues tested a petition versus a survey in an otherwise identical mailing for one of our clients. The table shows what happened.

Petition Versus Survey Test Results.

Pkg.	Total Gifts	% Resp.	Avg. Gift	Gross Rev.
Petition	295	0.74%	\$34.39	\$10,145
Survey	218	0.55	27.63	6,023
Difference with petition	+77	+0.19	+6.76	+4,122
% difference with petition	+35%	+35%	+24%	+68%

By our calculations, the difference in response between the two variations was clearly significant.

Who would have guessed this? Not I, certainly!

This just goes to show all over again that the market holds more mysteries than do our imaginations.

6

Tests or Rollouts?

It may just be the biggest problem in direct mail fundraising. And if you've conducted more than two or three donor acquisition mailings, you've probably already experienced it. Chances are, at least once, you've gotten *great* results from the initial test mailing of a new package only to see response take a nosedive on the rollout. And because you could *swear* you did everything right, there is no conceivable explanation for the discrepancy in results.

If results shot *up* from the test to the rollout, the discrepancy would be considerably less disconcerting, and it would hardly be the biggest problem in direct mail fundraising. But somehow it never seems to work out that way. All too often we're left red-faced,

The concepts of K. C. Blair discussed and quoted in this chapter are taken from Blair's article "'Shoddy Targeting' and the Disparity Between Test and Rollout Response Rates," *Journal of Direct Marketing*, Spring 1991, 5(2).

baffled—and considerably poorer—because direct mail doesn't work the way it's "supposed" to work.

Well, finally, after more than a dozen years in the industry, I came across a cogent explanation for this painful and expensive phenomenon. The explanation is *shoddy targeting*, a phrase coined by marketing consultant K. C. Blair. Blair says that "in about 63 percent of all direct marketing programs . . . when you think you know whom you're targeting, . . . you are actually targeting someone else."

Did you catch that? Blair says "63 percent"—nearly *two-thirds* of all mailings!

Worried yet? Wait. It gets worse.

Blair says he uncovered this phenomenon in the 1970s, when he started physically inspecting his lists and packages at the letter-shop. Addresses didn't match the targeted groups. Random selections weren't really random. Sometimes one list was drawn from zip codes starting with low numbers, and another was drawn from high-numbered zip codes. Lettershops also frequently reversed test and control groups.

According to Blair, at least one of these problems occurred in *half* of all his test mailings in the course of seventeen years, and about one in every four rollouts was marred because the lists actually used didn't match the mailer's specifications for the target group. This shoddy targeting affects response rates because "for every 100 tests conducted about 50 go to the wrong households. For every 100 rollouts there is a problem with 25."

If the problems Blair describes occur randomly, then the inaccurately targeted tests produce inaccurate comparisons 50 percent of the time, and the *accurately* targeted tests are paired with *inaccurate* rollouts one-quarter of the time, or in about 13 percent of

all the pairs. So, misleading response rates are present 63 percent of the time.

Why, though, when test and rollout results are divergent, does the rollout almost always seem to perform *worse* than the test rather than better? Blair says, “This is because it is difficult to find a successful combination of variables that test well, and any deviation from those success variables in a rollout will very likely result in a response rate that is less.”

Although the logic of this last argument strikes me as a little shoddy in its own right, Blair’s overall explanation for the frequent divergence between test and rollout results does ring true. In general my own experience in direct mail fundraising over roughly the same period of time that Blair was in the business suggests that the more tightly supervised and controlled the process of generating and affixing names and addresses, the more likely it was that I could trust a mailing’s results. In the more closely supervised cases, a knowledgeable and independent person usually conducted the sort of physical inspection of lists and mailing packages advocated by Blair.

That’s the *only* way to avoid shoddy targeting, he asserts, because “the phenomenon seems to be independent of the size or reputation of the list compiler and lettershop. All suppliers claim to have quality control functions, but we have found none that do it in a way that catches and corrects [these] problems.” The single biggest flaw in the system, Blair says, is that list selection criteria are misinterpreted by inadequately supervised “lower echelon people.”

I seriously doubt that “lower echelon people” are the problem. It seems more likely that the problems Blair describes lie in working with list brokers or computer service bureaus that are less

PERSPECTIVE: Pop Quiz

Attention, please. Listen carefully, and no fair peeking at your neighbor’s answer.

Let’s say you have just conducted a four-way test mailing totaling 100,000 names divided into four test panels of 25,000 names each. You tested three packages, each of which cost exactly the same amount, and one of the three was mailed to two panels. The table presents your results.

Four-Way Mailing Test Results.

Pkg.	Total Gifts	% Resp.	Avg. Gift	Gross Rev.
A, panel 1	284	1.14%	\$30.37	\$8,625
B, panel 2	243	0.97	27.88	6,775
A, panel 3	250	1.00	31.35	7,838
C, panel 4	290	1.16	27.04	7,842

than reliable. (Disreputable brokers have even been known to “stack” their lists on the front end with the very best and most current donors to inflate test results!)

Nonetheless, Blair made a major contribution with his analysis of shoddy targeting. His thesis comes the closest to explaining this age-old problem that so very often has undermined my own and other fundraisers’ faith in the validity of direct mail testing.

In these first six chapters I’ve introduced a number of the ambiguities and problems that make interpreting test results a challenge—and ongoing testing a necessity. The above perspective now gives you an opportunity to practice racking your brain over the issues presented by a typical set of test results and then making a rollout decision.

Now it's time to expand your donor acquisition program. Clearly, these results suggest that you can mail in larger quantities. But which package do you select as your control for a rollout mailing?

- Package A, because it generated the greatest revenue?
But, wait—did it really? Package A was mailed to both panels 1 and 3, and with panel 3, Package A yielded less revenue than Package C. So, how can you make any decisions about Package A?
- Package C, because it yielded the highest rate of response?
- Package A, because it consistently produced the highest average gift?
- And what, if anything, would you do with Package B?

Send your comments to me, at mal@malwarwick.com, and I'll respond with my own analysis of this mind-teasing puzzler.

Results Worth the Wait

Let's not kid ourselves: testing isn't advisable in every direct mail fundraising project. To conduct a meaningful direct mail test, you need the following four ingredients, any one of which may be in short supply:

1. *Money* to invest. Testing can be costly in the short run because it may raise printing and postage costs—and sometimes creative costs as well.
2. A relatively large mailing *quantity*. You will have to have a large list (or lists) so you can extract test panels large enough to provide statistically significant results.
3. A *commitment* to retesting. Solitary test results should be swallowed only with a grain of salt.

4. Plenty of *time* to evaluate results. Responses from competing test panels might be received at different rates.

Time may be the *most* important of these four ingredients. Far too often direct mail fundraisers jump to conclusions on the basis of early test results. As often as not those early conclusions are *wrong*.

To show just how far off target you can stray with a premature reading of test results, consider the experience of one nationwide women's organization. The organization's direct mail program had reached an impasse because a long-running acquisition control package was no longer working well. The group's donor base was starting to shrink as a result. However, a bold new test package raised hopes for a revival. With fingers crossed all around the organization, 46,000 pieces of the new package were dropped.

Then nothing happened.

Or to be more precise, *almost* nothing.

After one week—five business days of returns—98 responses had been received, a mere 0.21 percent.

Three harrowing weeks later, by the twentieth business day of returns, the response rate stood at only 0.58 percent. This was well past the *doubling point* in the group's prior acquisition mailings (that is, the point at which half of each mailing's total revenue had been received).

For an organization accustomed to acquisition response rates in excess of 1.50 percent, this was devastating news. Nearly a month after the first significant returns, it had recovered less than a third of the mailing's costs and attracted a mere 268 new donors to support its work. At comparable points in the past, the group

had usually received the lion's share of the revenue from acquisition mailings and recovered between 70 and 80 percent of costs.

Fortunately, common sense prevailed. Instead of acting precipitously on the basis of these discouraging early results, the organization resigned itself to waiting.

Patience saved the day.

This acquisition test mailing eventually yielded a slender profit—*after nearly eight months of results!* More important, the ultimate response rate of 1.60 percent—and the average gift of more than \$27—conclusively proved the organization's continuing appeal to prospective donors and established a viable new control package. (More detailed figures are displayed in Table 7.1.)

All it took was time.

Sometimes it really *pays* to wait. And the converse is equally important: *not waiting* can lead to big mistakes!

Of course, while you're waiting, you are likely to start second-guessing every aspect of your package. But it's always wise to test before you leap, as this chapter's perspective shows.

TABLE 7.1. Eight Months of Revenue Receipts in a Donor Acquisition Test

# Weekdays	Total Gifts	% Resp.	Avg. Gift	Gross Rev.	% of Gross
5	98	0.21%	\$25.01	\$ 2,451	12%
20	268	0.58	22.93	6,145	31
40	565	1.23	25.65	14,492	72
55	632	1.37	26.85	16,969	84
80	714	1.55	26.95	19,241	96
172	739	1.61	27.25	20,141	100

PERSPECTIVE: Write Long or Short?

How many times have you wondered whether the public’s patience with long direct mail letters has finally, irretrievably worn out? Certainly, with all the talk about baby boomers’ fast-evolving attitudes and unpredictable behavior, I’ve wondered whether the old rules of direct mail fundraising still apply. More to the point, although it is an article of faith in direct marketing that “long copy pulls,” not a month goes by without one client or prospective client protesting how irrational that sounds.

This has led me and my colleagues to test—and retest repeatedly—the old saw that “long copy outpulls short.” From time to time the results are equivocal. But as best I can tell, it’s not wise to buck the conventional wisdom, at least in this case.

More often than not, the old, familiar four-page letter generates higher response than an abbreviated, two-page version. The table presents the latest data I’ve seen on the frontlines of testing:

Long Letter Versus Short Letter Test Results.

Pkg.	Total Gifts	Avg. Gift*	Gross Rev.*
Short letter	229	\$21.38	\$4,725
Long letter	310	22.46	6,649
Difference with long letter	+81	+1.08	+1,924
% difference with long letter	+35%	+5%	+41%

* These figures take into account only gifts of less than \$100.

The conclusion? Try new forms and formats. Keep testing—constantly. But don’t make the mistake of assuming that instinct is better than testing. It just ain’t so!

Reading Your Results

Now, here comes the tough part. Consider two hypothetical direct mail tests and their results. In Mailing A, 10,000 letters were mailed in *window* envelopes and 10,000 in *closed-face* envelopes. The window envelopes generated 500 gifts, and the closed-face envelopes yielded 525. In Mailing B, 1,000 letters were mailed with brochures and 1,000 identical letters were mailed without brochures. The brochure packages produced 5 gifts, the no-brochure packages 20 gifts.

What conclusions can you draw from these two sets of test results? Is it true that in Mailing A, the closed-face envelope significantly outperformed the window envelope? After all, both versions of the package generated hundreds of gifts—surely, more than enough to establish credibility! And is it also true that the tiny numbers of gifts in Mailing B call into question the validity of those test results?

The answer to both these questions will be a resounding no! if you ask Tom Collins. Tom Collins's name isn't widely known among fundraisers. But to those direct mail fundraisers who are familiar with the wider world of direct marketing, Tom Collins is a legend. He was the cofounder of Rapp & Collins (now Rapp Collins Worldwide, part of the massive, international Omnicom Group), one of the giants of the advertising industry. And he is the coauthor (with Stan Rapp) of two immensely influential books, *Maxi-Marketing* and *The Great Marketing Turnaround*. He's also one of the all-time great copywriters—and one of the nicest guys you could ever meet. And by the way, he's not exactly a stranger to fundraising direct mail. Collins's letters for the 1972 McGovern presidential bid netted, in today's dollars, at least \$6 million before and during the primaries and \$30 million during the fall election campaign.

So listen closely to what Tom Collins would say in response to the two questions I posed above. In Mailing A, the difference in results between the closed-face and window envelopes is too small to matter. It's *not* statistically significant. In contrast, in Mailing B, the difference between the brochure and no-brochure packages is meaningful—despite the small absolute numbers.

How can this be? Don't these assertions fly in the face of all logic? Here's how Collins explained this perplexing matter in a long, thoughtful letter he recently wrote to me:

Significance is not determined by size of response per se. The larger the percentage difference, the smaller the total can be in order to be statistically significant.

For example, if you received 5 responses from Copy A and 20 responses from Copy B, I am sure that would check out as a statistically significant difference, even though the total of 25 is very small. Under the Law of Probability, if you flipped a coin 25 times

over and over again, it would take many, many tries before it finally came up 20 heads and 5 tails, or vice versa.

Conversely, of course, the smaller the percentage difference, the larger the total must be to reach statistical significance. Thus, a difference of 15 responses is very meaningful in the coin-toss example cited above; it would be meaningless in a total response of 1,000.

And since subtle changes produce small differences, the more subtle the change, the bigger the test cells that are required.

The old rule of thumb we were taught was pretty handy—that in a 2-way split, any difference which is more than double the square root of the total has a 95% chance of being valid. Thus, if you got 60 responses from Copy A and 40 from Copy B for a total of 100, a significant difference would be 10 times 2, or 20, so the result would be on the borderline of significance.

Let's apply Collins's formula to Mailings A and B. In Mailing A the total number of responses received from both mailings was 500 plus 525, or 1,025. The square root of 1,025 is 32. For a 95 percent chance that the difference in results between the closed-face and window envelope packages is *significant*, that difference would have to be greater than 32 times 2, or 64. Because the actual difference was 25, the results of the test in Mailing A are not significant.

If the actual difference had been, say, 50 instead of 25, the likelihood that the test had produced meaningful results would be somewhat greater—but still a far cry from the 95 percent confidence level that *I* would like to see if I had to make an expensive production decision.

Now let's take a look at Mailing B through the lens of Tom Collins's formula. Here, the sum of the responses to both package versions is 25, of which the square root is 5. Twice 5 is 10—a number comfortably smaller than the difference between the returns reported for the two test panels. In other words, following the

square root formula, we can conclude that omitting the brochure from our packages will significantly improve our mailing results.

Well, maybe. Here's where statistics and common sense may be in conflict once again. Although I wouldn't hesitate to follow Collins's formula in interpreting the results of Mailing A, I'm inclined to take it with a grain of salt when it comes to Mailing B. With the absolute numbers so very small—and so many slipups possible at the lettershop as well as in counting the returns—remember shoddy targeting?—I would insist on *retesting* results that rest on such a narrow base. In this and similar cases, comparing the difference in results to the square root of the sum can be taken as a useful indication but not a completely conclusive one.

It's easy for anyone to use the formula Tom Collins describes. And there's a very simple reason it's important to go to all this trouble. As Collins wrote to me, "Wishful thinking frequently enters into interpretation of results. Thus, if one wants Copy A to win over Copy B, then one will crow with delight if it does, regardless of the numbers. But if one wants Copy B to win, one will say, '[Copy A pulling better than Copy B] doesn't mean anything, the numbers are too small.'"

Because testing—and direct mail fundraising in general—is about real-world *facts*, not fantasy, we would all be well advised to apply Tom Collins's simple formula whenever we evaluate testing results.

Not to do so could cost your organization a bundle! And a related pitfall to watch out for is described in the following perspective, which shows that large gifts can affect the statistical significance of your results.

In the following chapter my colleague Dan Suzio takes a closer look at the concept of statistical significance.

PERSPECTIVE: What Counts?

Consider the results of this test for one of my firm's clients. We mailed each of two different packages—costing the same but using different copy—to 40,000 names randomly selected from the same merged group of lists. The figures in the first table show what happened overall.

Test Results Including All Gifts.

Pkg.	Total Gifts	% Resp.	Avg. Gift	Gross Rev.
A	172	0.43%	\$26.47	\$4,552
B	226	0.56	22.89	5,173
Difference with B	+54	+0.13	−3.58	+621
% difference with B	+31%	+30%	−16%	+14%

Common sense says that package B outperformed package A; it brought in more gifts (and thus more money). Some fundraisers, preferring the higher average gift achieved by package A—because bigger average gifts imply greater donor loyalty and generosity—would fasten on that big difference and call it a wash, possibly even preferring to remail A rather than B. But that would be wrong. The flaw in this approach is that the mailing yielded a total of seventeen gifts of \$100 or more, which together were capable of distorting the overall results. In this mailing the number seventeen is much too small to be statistically significant. It was purely a matter of chance that one of the packages generated more such big gifts than the other did.

The second table shows what the picture looks like when those seventeen \$100+ gifts are factored out of the calculations (in all the columns, not only the average gift and revenue columns).

Test Results Excluding \$100 Gifts.

Pkg.	Gifts*	% Resp.*	Avg. Gift*	Gross Rev.*
A	160	0.40%	\$20.80	\$3,328
B	221	0.55	21.15	4,673
Difference with B	+61	+0.15	+0.35	+1,345
% difference with B	+38%	+38%	+2%	+40%

* These figures take into account only gifts of less than \$100.

This recalculation reveals that package A accounted for twelve of the seventeen large gifts. Excluding them, the disparity in revenue between the two packages is even greater—and the average gift is almost identical. So it would be a big mistake to elect package A over B!

Tricky business, no?

Understanding Confidence Levels

Dan Suzio

We often hear the term *statistically significant* when discussing opinion polls, medical research, or direct mail test results. But what do those words really mean? Is statistical significance an unambiguous, yes-or-no question? Is there a simple and accurate way you can measure it? And if there is, does that mean you can ignore any test results that are shown to be *insignificant* and just roll out your *significant* tests in ever larger quantities?

Let's say you've just tested a new acquisition package. Your control package focused on asking donors to support a good-government organization, and your test package asked donors to resist the efforts of people who oppose the organization. You mailed 50,000

Special thanks to Jeff Frost for his mathematical expertise and patient explanations.

TABLE 9.1. Hypothetical Initial Test Results

Pkg.	# Mailed	Total Gifts	% Resp.
Control, “help the good guys”	50,000	321	0.64%
Test, “beat the bad guys”	50,000	367	0.73

of the control to one group and 50,000 of the test to another. You worked with your data processing house to make sure the groups were as similar as possible—drawn equally from the same lists and distributed equally to the same zip codes—and you made sure your lettershop delivered all 100,000 letters to the post office at the same time. Now you’re looking at the results, as shown in Table 9.1.

Obviously, the test package is the winner. Or is it? We can see it generated forty-six more gifts than the control, but we all know from past experience that random (or at least unexplained) variations can occur in response rates, even when we mail the same package to the same lists.

How can we tell whether these results are valid and repeatable—or simply an example of random variation?

The Definition

To answer that question, statisticians have invented something called the *confidence level*. Simply put, the confidence level is the likelihood—expressed as a percentage—that the results of a test are real and repeatable and not just random. The idea of the confidence level is based on the concept of the *normal distribution curve*, which shows that variation in almost any data (such as the heights of all fourth graders or the amount of rainfall in January)

tends to cluster around an average value, with relatively few individual measurements at the extremes.

Using a formula (presented later in the chapter) we can calculate that the difference in response rates in our example has a confidence level of 92 percent. This means that according to probability theory, there's a 92 percent chance that you would see similar results in a repeat of the test. (It does *not* mean you would receive the same number of gifts or that the difference between the responses to the packages would be the same—only that as in the original mailing, the test package would be likely to receive *more* gifts than the control.)

Once you know the confidence level for your results, the next question is whether that level can be considered statistically significant. For example, is our 92 percent confidence level significant?

A confidence level of 50 percent means the difference is truly random, with only a 50-50 chance that you would see the same results in a repeat of the test. Even at 75 percent the odds are not good: there's a one in four chance that your results are meaningless. Some statisticians consider 90 percent the minimum confidence level for statistically significant results, and that's reportedly the standard used in many election polls. Others insist 95 percent is the lowest level that can be considered significant. In medical research, for obvious reasons, there's a strong preference for even higher levels of confidence.

So the results in our example are—well, they're OK. Not perfect, but not exactly meaningless either. Some fundraisers would declare a winner and move on, whereas the more cautious would want to retest. Still others would examine the results for each list used in the mailing, looking for patterns. Personally, I'd go with a retest, using a larger quantity if possible.

It's important to remember that we're talking about probabilities, and there's no magic number that guarantees your results will be repeatable. Although it's always best to have a confidence level of 95 percent or higher, you shouldn't ignore results in the 80 to 90 percent range. Those results may indicate trends and provide clues about how to improve your mailings; at the very least they're worth retesting, preferably in larger quantities. (In any test, a larger sample size will generally give more reliable results.)

The On-Line Calculator

Because you've been interested enough to read this far, you deserve a reward. Here it is: the Confidence Level Calculator. This useful tool is located on the Mal Warwick & Associates Web site, at <http://www.malwarwick.com/conflevel.html>. All you need to do is enter the quantities you mailed (omitting all commas) and the number of gifts you received. The calculator does the rest.

The perspective for this chapter gives another example of considering statistical significance.

The Formula

This is the geeky part. If you insist on determining confidence levels yourself, here's what you need to know. (But I have to warn you, it's a lot easier to use the Mal Warwick Web site.)

There are some relatively simple formulas for determining whether a direct mail test is statistically significant, including the one featured in the previous chapter. The problem, though, is that they generally give only a yes or no answer—and any statistician will tell you that probability theory never produces results that are

PERSPECTIVE: Significant?

In a large prospect mailing, three test panels of 30,000 names each were set aside to compare the impact of suggesting a different minimum gift (both in the letter and on the response device). The decision to conduct this test was shrewd, because the ask amount is more likely than most factors to affect a mailing's response rate. What happened is displayed in the table.

Ask Test Results.

Pkg.	Total Gifts	Avg. Gift*	Gross Rev.*
A, \$20.00 ask	223	\$18.61	\$4,150
B, \$21.50 ask	207	\$20.34	4,210
C, \$35.00 ask	206	\$21.11	4,349
Difference between B and A	+16	+1.73	+60
Difference between C and A	+17	+2.50	+199

* These figures take into account only gifts of less than \$100.

The differences in response among the three panels *are not* statistically significant. When we apply the rule-of-thumb formula discussed in Chapter Eight, we see that the difference between the responses to Package A and Package B would have to be greater than 41 to be significant and that the same holds true for the difference between the responses to Package A and Package C, whereas the actual differences are much less. If we ascertain the confidence levels for these tests, we find that the level for the results of both the Package A to Package B comparison and the Package A to Package C comparison is less than 80 percent. Therefore, dramatically raising the ask amount did not meaningfully depress response.

Even more interesting, the differences in the ask did not yield substantial discrepancies in the resulting average gift or in the gross revenue!

Be warned, however; if your organization conducts such a test, you can reasonably expect the results to be different from these. *Do not* conclude that the ask amount makes little difference. A few months earlier, a somewhat similar test for the same organization—with different packages—produced contrary results. And as this book has pointed out earlier, testing the ask is typically a high-potential test. The moral of this story? Don't jump to conclusions when reading test results. Always analyze them carefully.

simply black or white. To understand your test results, you need a relatively precise number for the likelihood that your results are repeatable—and at the same time, you need to understand that this number is only a probability, not a guarantee.

The first step in determining confidence level uses a formula to find something called the z value; once you know that value, the next step is just to look up the corresponding confidence level. The formula may seem intimidating at first, but it's just basic algebra.

First, write down the quantity mailed (N_1 and N_2) for each of your two test panels. Using the figures from our earlier example, we get

$$N_1 = 50,000$$

$$N_2 = 50,000$$

Next, figure the response rate for each test panel (P_1 and P_2), expressed as a decimal (not a percentage):

$$P_1 = \frac{321}{50,000} = 0.00642$$

$$P_2 = \frac{367}{50,000} = 0.00734$$

Now, substitute the response rate numbers for the variables in this equation:

$$z = \frac{P_1 - P_2}{\sqrt{\frac{P_1 (1 - P_1)}{N_1 - 1} + \frac{P_2 (1 - P_2)}{N_2 - 1}}}$$

If you're trying to solve this equation on your desktop calculator, you'll discover that given typical direct mail quantities, the numbers inside the square root sign are extremely small. If your calculator can't handle ten places to the right of the decimal point, it will probably round the whole thing off to zero. Fortunately, there are alternatives. Most scientific calculators include a function for determining z values, so you could just ask a high school algebra student to solve the equation for you. (They get this sort of thing as homework all the time.) Or you could solve it in one step using Microsoft Excel or any other spreadsheet program, substituting cell references for the variables. In that case, you'll want to write the equation this way:

$$z = (P_1 - P_2) / \text{sqrt}(((P_1 * (1 - P_1)) / (N_1 - 1)) + ((P_2 * (1 - P_2)) / (N_2 - 1)))$$

Depending on which test panel generates more gifts, z may be a positive or a negative number; if it's negative, ignore the minus sign. (In other words, use the absolute value of z , or $|z|$.) Now, all you have to do is look up the confidence level on the following chart:

<i>If z is at least . . .</i>	<i>The confidence level is . . .</i>
1.280	80%
1.310	81
1.340	82
1.370	83
1.410	84
1.440	85
1.480	86
1.515	87
1.555	88
1.600	89
1.645	90
1.695	91
1.750	92
1.810	93
1.880	94
1.960	95
2.055	96
2.170	97
2.330	98
2.575	99

Using our example, we find that $z = -1.760$, and the confidence level is 92 percent.

Finally, a note for the fundraiser who may occasionally encounter scientific or technical publications: most scientists and other researchers don't refer to confidence level in their published work. Instead, they use the p value, which is expressed as a decimal and is "the evidence against rejecting the null hypothesis." So, if a scientist were reporting the example used here, he or she would be more likely to write " $p = 0.08$ " than "the confidence level is 92%." For the rest of us, though, it is often habitual to think of a higher number as better, and expressing this number as a percentage of confidence—with 100 percent being the best (although unattainable) result—feels like a result we can do practical work with. So the p value is subtracted from 1 to obtain the confidence level.

A few words of caution about all of this. These confidence levels are valid only when you're comparing test panels that can be thought of as existing within a single event. Don't try to adapt the formula to a situation that changes over time (such as your total number of active donors) or to use it (or the on-line calculator) to compare appeals that mailed at different times (there are so many uncontrolled variables in that case that the comparison may be meaningless). You'll notice that the formula does allow you to use test panels of different sizes, but when you do, make sure your merge-purge house is extremely careful about producing statistically equivalent lists of names.

You should also keep in mind that this chapter addresses only the question of response rate—not average gift, donor acquisition cost, or other factors. Evaluating the statistical relevance of those variables requires other, more complicated, formulas.

Finally, as with any other mailing, you'll also need to evaluate the results of your tests in terms of your overall fundraising goals.

The Value of Guesswork

Let's play hunches. Which of the following two changes in a direct mail package is likely to improve fundraising results? Or are both likely to do so?

- Using a 9-by-12-inch outer envelope, or *carrier*, instead of a 6-by-9-inch carrier.
- Affixing live postage stamps to the reply envelope instead of using a business reply envelope (BRE).

One East Coast charity tested both these questions in a mailing to its recent donors a while back. And—as happens so often in this unpredictable business—the results surprised the fundraisers. Their hunch didn't prove true.

The charity's working hypothesis was that *both* the 9-by-12-inch envelope *and* the live return postage would improve results in the mail.

Not so.

The mailing was structured as a three-way test. One group of donors received a fundraising appeal in a 6-by-9-inch package with a BRE. A statistically identical group of donors was mailed the same appeal in a 9-by-12-inch package with a BRE. A third group received a 9-by-12-inch package that was identical in every respect save one: the reply envelope bore a 25-cent stamp (which was good enough for the U.S. Postal Service a decade or so ago).

Despite industry scuttlebutt that large envelopes attract greater attention and thus increase response, the 9-by-12-inch envelope in this mailing proved to be a bust. The oversize carrier actually decreased response. And because it cost much more to print than the 6-by-9-inch carrier, the larger envelope had an even more pronounced, negative effect on net revenue. Had this charity mailed all its appeals in 9-by-12-inch envelopes instead of 6-by-9-inchers, it would have lost many tens of thousands of dollars.

So much for industry scuttlebutt—in this case at least.

But the live return postage proved another matter altogether. The stamped reply envelope *lifted* response by between 22 and 41 percent over the BRE (with the rate varying from one donor file segment to another). Even discounted for the substantial additional cost of the postage, net revenue jumped by between 15 and 30 percent. And, interestingly, the lift was more pronounced in file segments with 1989 donors than it was in segments with the 1990 donors.

But if you feel tempted to draw global conclusions from this organization's experiment, be careful—what has worked for one charity won't necessarily work for you. Time and again, direct mail

fundraisers have stumbled when trying to translate successful results for one mailer into programs for other mailers. Indeed, this test proves that point, because the industry scuttlebutt that spawned the idea for this test must have grown out of *somebody's* real-world experience. Yet only one of the two hunches tested lived up to expectations.

Hunches, industry scuttlebutt, and even the examples in this book may be excellent *starting points* for your own direct mail fundraising tests. But always be certain you observe rigorous testing procedures—because you'll never know in advance exactly how your test will turn out.

Moreover, some hunches and conventional wisdom may have more potential than others. As the following perspective shows, and as I have been emphasizing throughout, the ask is often a great choice for your research project.

PERSPECTIVE: Once Again, How Much?

Direct mail veterans know that few factors in a mailing will influence the response rate more than price—or in the case of a fundraising appeal, the ask. And testing the limits of the ask can pay off in a very big way, as we saw in the perspective for Chapter Three.

With visions of sugarplums in mind, a client recently tested its control ask of \$35 against a specially sweetened \$100 ask. The mailings were sent to two test panels of about 18,800 names each. Except for the suggested gift amount, the two packages were identical in all ways but one: the \$100 ask was featured in a small insert illustrating the attractive premium donors would receive if they complied with the request. The table tells the rest of the story.

More Ask Test Results.

Pkg.	Total Gifts	% Resp.	Avg. Gift	Gross Rev.
Control, \$35 ask	124	0.66%	\$54.15	\$6,715
Test, \$100 ask	72	0.38	65.83	4,740
Difference with \$35 ask	+52	+0.28	−11.68	+1,965
% difference with \$35 ask	+72%	+74%	−22%	+41%

Clearly, the \$100 ask was a loser—big-time. Not just because the \$35 ask pulled almost twice as many responses but also because the difference in average gifts was comparatively slight. The upshot was that the client received about \$2,000 less in revenue because it mailed this test to half the selected prospects rather than mailing its control to all the prospects.

Once again some lessons are reinforced: first, testing confirms that the ask amount matters, and second, testing can be an expensive proposition. And there's yet another lesson in the background—only by testing can you learn how donors will really respond.

Testing Tips from Major Mailers

When we direct mail fundraisers get together, we talk about *testing*. Because when you get right down to cases, testing is all we've got to talk about. Or so it seemed, at any rate, at a fundraising forum held in New York a while back by Moore Response Marketing Services. Twelve major direct mail fundraisers were represented, among them UNICEF, the National Jewish Center for Immunology, Father Flanagan's Boys' Home, the American Red Cross, CARE, the Smithsonian Institution, and the Society for the Right to Die. A report on the discussion came to me shortly after the session, and it proved all over again that direct mail fundraisers shouldn't take *anything* for granted.

For example, one major mailer reported improved results from eliminating from marginal lists the names that were insufficient in number to qualify for extra postal discounts: for this

mailer these names had a *lower* response rate than names that did qualify for these postal discounts. The procedure reduced the organization's mailing costs and boosted the overall response rate. However, another mailer reported the *opposite* experience: the basic/residual-sorted names were *very* responsive for this mailer. Eliminating them from its mailings would have caused a significant revenue loss.

What caused this difference? One possible explanation identified by forum participants was that the first mailer worked primarily in *urban* areas, with names concentrated in relatively few zip codes, whereas the second mailer relied on more widely scattered *rural* respondents. In other words, these reports proved once again that situations are never identical; what works for one organization may not work for another—and blindly following another's test results may be dangerous to your fiscal health.

Here are more highlights from the forum:

- One experienced mailer advised caution in the headlong rush to downsize packages now classified as *flats* under U.S. Postal Service rules (because they exceed the dimensions the USPS favors). Despite their high postage rate, flats sometimes still pay for themselves by yielding higher response (perhaps because they face less competition and stand out more in the mail). Another mailer found a second reason not to abandon flats: for this organization they yielded higher-dollar donors, who have proven better lifetime givers for that organization—suggesting that the added investment in flats will pay off in the long run.
- A merge-purge bureau told of the time it had suggested deleting names of deceased persons from a customer's file, by check-

ing that file against an updated proprietary file (presumably, a *file of the dead*). But the customer rekeyed and mailed those names “just one more time”—and the results were equal to or better than those from the rest of the mailing! (I do *not* recommend this procedure. There are few surer ways to discourage legacy gifts than alienating a widow or widower.)

- A large charity with many local chapters said it had tested whether donors would respond better to direct mail appeals identified as coming from a local chapter rather than from the national organization. *There was no discernible difference.*
- One participant wisely urged large mailers to track donors back to the original lists from which they came and the original packages that attracted their first gifts. Five years of testing showed that lists yielding higher average gifts (and lower response rates) were a better long-term investment than lists producing lower average gifts (and higher response rates).
- One mailer had experimented with recyclable paper and inks and had found that both can make a difference in response. Unfortunately, it wasn't the response environmentalists would have wanted: both changes *decreased* response. But another organization reported *no meaningful difference* as a result of using recyclable materials.

Of course you would need to test each of these results for yourself before acting on any of them in a big way, even if, as the following perspective demonstrates, all you sometimes learn is that you could have taken something for granted after all.

PERSPECTIVE: Response Devices?

One element of the conventional wisdom about direct mail fund-raising is that a detached response device written and designed to match the style and substance of an appeal is almost certain to out-pull the response to either a tear-off “coupon” or a generic, wallet-flap reply envelope. Well, at least one recent test demonstrates that the conventional wisdom is . . . indeed sometimes wise.

A long-standing control package for one of our clients included a tear-off response device. Despite the added cost (nearly \$0.03 per package), we advised the client to substitute a stand-alone form, mailing 30,000 packages to each of two test panels. What happened is summed up in the table (which in addition to the usual categories also shows the acquisition cost per donor [Acq. Cost/Donor]).

Response Device Test Results.

Pkg.	Total Gifts	% Resp.	Total Cost	Acq. Cost/Donor*
Control, integrated response form	216	0.72%	\$10,560	\$16
Test, stand-alone response form	274	0.91	11,430	17
Difference with stand-alone	+58	+0.19	+870	+1
% difference with stand-alone	+27%	+26%	+8%	+6%

* Acquisition cost per donor = total cost – gross revenue / number of gifts (responses).

“Aha!” you say. “It’s true that the stand-alone form resulted in a 27% higher response rate—but because of the added printing cost, the acquisition cost actually rose by \$1 per new member!”

True enough. But that slight elevation in the acquisition cost was far outweighed by the substantially greater number of new members acquired. And after all, acquisition is (usually) all about broadening the base of support.

In addition to my usual cautions about the fact that a similar test for another organization under other circumstances might yield very different results, I must also point out here that in house-file mailings, my colleagues and I frequently use other response device formats than the stand-alone format, for various reasons.

Testing Is Fundamental

Some time ago the newsletter *What's Working in DM & Fulfillment* interviewed five direct marketing professionals on the subject of testing: Michael Edmondson (*American Demographics* magazine), Marc Jacob (Bantam Doubleday Dell), Vince Dema (Vince Dema & Company), John Klinge (*Hippocrates* magazine), and fundraiser Kay Partney Lautman (Oram Group Marketing). Their observations are as true today as they were then, and so I offer them here. In adapting them for this chapter, I have, however, translated some of the original terms and examples into those more familiar to fundraisers

If you're looking for hard and fast rules on testing, you'll find instead that there are about as many rules as there are direct mail

The interview on which this chapter draws appeared originally in "Testing Tips," *What's Working in DM & Fulfillment*, August 28, 1989.

fundraisers. But even though there are no right or wrong answers across the board, there are certain factors you should always consider. So, before you begin testing your direct mail or telephone fundraising fulfillment packages, heed these five fundamental guidelines from the pros:

1. *Look at your long-term fundraising goals.* Then integrate your testing goals into that long-term planning. In one direct mail fundraising program, for example, the long-term goal might be to acquire as many donors as possible, regardless of the size of the average gift. Another organization's strategy might be to acquire new donors at the highest feasible initial gift level and then to concentrate on upgrading those donors. Such contrasting strategies call for very different approaches to testing, focusing on response in the former case and on the gift level in the other.

Do you want to acquire new donors at a lower cost? Then test suggested minimum gift levels. Do you want larger numbers of donors? Test new packages or package components that might dramatically boost response.

2. *Control possible losses.* To make sure you get the overall response necessary for a successful mailing, use the historical response rate of the control package to determine the quantity you need to mail to reach that goal. Send the control to at least that many names, and select those names from your proven lists. For example, if you need 1,000 donors and your control historically pulls 1 percent, send the control to at least 100,000 names from proven lists. Then, for a valid test of one or more different packages, mail the test package(s) to a random selection of names who are on the same lists but who have not been sent the control package.

Fundraisers use various rules of thumb for determining the quantity mailed in each test panel, or cell. It's common to set aside

between a fifth and a third of the names in an acquisition mailing, and divide those names into test cells of between 10,000 and 30,000 names each. (Remember: you need at least two of these cells for a valid test: one for the control, the other for the test.) The remaining names, normally the bulk of the mailing, constitute the third, or *remainders*, panel.

3. *Look at the results of past tests.* Before starting a new test, check previous results for possible trends. Test history is a valuable tool for identifying package or list fatigue. For example, you may notice certain proven lists not pulling as well over time even when you are mailing the control. It's important to determine whether that's due to mailing too often to the list or to a package that needs to be changed.

Combined results of past tests may show that the test you are considering may not be as valuable as another test. For instance, if you're thinking of testing a *lift note* (typically, a signed note or short letter that reinforces the case for giving) but see that all previous lift note tests mailed to your market have never made more than a 0.1 percent difference, another lift note test will most likely not make a vast improvement in your response. Conversely, if a teaser copy test and a sticker test have both outpulled any lift note test, and the sticker became part of the control package because it pulled best, consider retesting teaser copy. In such circumstances, it's always more likely than a lift note to affect the results significantly.

Another reason to look at past tests is to determine whether their results are in fact valid. For example, if a test of a premium bombed, check the package to see if the premium was promoted properly or if it was buried somewhere in the package.

You might also consider whether changes in donors' attitudes might have occurred and might justify retesting, as the following perspective discusses.

PERSPECTIVE: Credit Cards?

For many years the conventional wisdom in direct mail fundraising held that offering donors the option of contributing by credit card would depress response. On several occasions in the 1980s I tested this proposition for various clients. Sure enough, the conventional wisdom had it right.

But have attitudes about credit cards changed, and with them the pattern of responses in direct mail appeals? There's mounting evidence that that may be the case.

The table, for example, offers a brief look at the results of a recent test my colleagues and I conducted in a member acquisition mailing for one of our clients, mailing 60,000 packages to a control group and to a test group.

Credit Card Option Test Results.

Pkg.	% Resp.	Avg. Gift*
No credit card	0.69%	\$18.84
Credit card	0.65	19.41
Difference with credit card	−0.04	+0.57
% difference with credit card	−6%	+3%

* These figures take into account only gifts of less than \$100.

The variation in response to the two offers is negligible. In other words, the credit card option did not depress response.

Will that always be the case? We've occasionally found the contrary with other clients and other packages. But the weight of evidence is shifting in favor of offering credit cards. Is it possible that many people are looking for more ways to use their credit cards, in order to accumulate various credits, such as airline mileage credits?

4. *Estimate response before you test.* Another way to decide that a proposed test is genuinely worthwhile is to estimate its potential response. If you know the response generated in the past by similar offers or package components or by using the same list, you can estimate the best response you're likely to get. Then decide whether that response is enough to make the test profitable.

5. *Base your analysis on your goals.* After final results are in, analyze them in relation to your fundraising goals. For example, if having a certain number of active donors is the goal, you will probably want to attach more importance to the net response and the extent to which your acquisition mailing moved you closer to your goal and less importance to the acquisition cost per new donor. Base your analysis on projected rollout quantities to get the bigger picture. Be sure to include all costs, especially those sometimes forgotten ones like postal reply charges.

Comparing Costs

Testing doesn't necessarily make sense for every nonprofit direct mail fundraising program. Sometimes mailing quantities or budgets are just too small to justify the added expense. Occasionally, a test may even be a downright bad idea. Whenever I'm in doubt, I compare the cost of conducting a test with the projected increase in revenue in the long run if the test should work. If the long-term value of the improvement isn't a whole lot greater than the short-term cost, I rule out the proposed test.

For example, experience tells me that a new letter-signer probably won't improve fundraising appeal results very much—perhaps 5 percent or less. So, if an organization expects to raise \$50,000 during the next year from successive remailings of the letter in question, it's not reasonable to expect more than \$2,500 in added revenue from a change in signers, and it may well be less.

There's no way to know for sure without testing a new signer against the old one, but that test won't come cheap. The organization will probably need to mail *at least* 10,000 letters bearing each signature—20,000 in all—and that will mean added printing and production costs all down the line, from the design shop to the lettershop. The test may entail additional postage too, because each test panel will probably have to be addressed in its own zip code sequence, and postal discounts may be lost in the process. All told, these additional costs may total \$2,500 or more.

In other words, I would leave this signature test on the drawing board. In all likelihood, most small nonprofit mailers would arrive at the same conclusion.

What to Test

Once you're clear that a test is well advised, it's time to decide what to test. One rule of thumb is to start your testing with the widest parameters you can in order to determine the potential range of possible response. For example, in a membership renewal series (a sequence of several dues renewal notices), you might run a gift upgrade test to see the results of requesting larger sums of money. If the average gift increases substantially but response falls off when you ask for a renewal gift twice last year's gift, you could then test asking for up to 50 percent more than last year's gift, then 30 percent more, then 20 percent. In each stage, you'll have something relevant with which to compare the results. (In contrast, testing a 15 percent upgrade against a 20 percent upgrade may gain you little or nothing in terms of either increased response or useful information.)

Most direct mail fundraisers will be wise to test the basics like the minimum suggested *gift* (the ask), the *offer* (including mem-

bership benefits and donor recognition), and package *format* (the kinds, design, and number of materials in the package, even, as the following perspective discusses, the typeface). But you have almost limitless possibilities for additional tests. Following are five test categories that organizations often explore:

1. *The ask*. What is the optimal contribution level you can request? For example, in acquisition mailings, if you currently ask for a minimum gift of \$25, will results improve if you specify \$30 or \$35?
2. *Postage rate*. If you're now mailing to your previous donors via third-class bulk mail, will first-class postage improve your net returns? In acquisition mailings, the answer is likely to be no, but in *house mailings*, using first-class postage with some segments may be very fruitful.
3. *Involvement devices*. Will a survey or other involvement device increase response to your donor acquisition package or annual appeal? Asking prospects or donors to perform some appropriate action in addition to—or sometimes instead of—sending a gift is likely to lift response. (It may, however, also lower the average contribution.) A typical action is signing a petition or postcard to a public official or agency.
4. *Premiums*. Is there a free gift you can offer donors that will elicit greater response to an acquisition package or annual appeal? Such free gifts as stamps, stickers, or decals are a mainstay of many successful, long-running direct mail fundraising programs. They may increase net revenue for your program too.
5. *Letter length and types of inserts*. Fundraisers traditionally mail four-page acquisition letters, but would you save

PERSPECTIVE: Courier Type?

OK, I admit it. I'm probably as guilty as anyone of preaching the gospel that fundraising letters should always be set in 12-point Courier (typewriter-like) type. It's one of the few points on which I've been dogmatic for much of my career. But perhaps I should have heeded my own advice and spurned dogmatism in this instance as well as in every other one. In any case I'm coming to feel more and more uneasy about categorically recommending Courier type for fundraising appeals.

In recent years my colleagues and I have conducted frequent mailings in which we test 12-point Courier head-to-head with 12-point Times Roman, the other (now standard) typeface in direct mail fundraising. And more and more frequently the results have looked like the ones from our most recent mailing, to two 60,000-name panels, as displayed in the table.

Typeface Test Results.

Pkg.	Total Gifts	% Resp.	Gross Rev.
Times Roman	412	0.68%	\$8,022
Courier	396	0.66	7,427
% difference with Times Roman	+4%	+3%	+8%

Will these results hold true in the case of every fundraising appeal? Of course not. I can certainly envision many appeals in which Courier could still be essential: for example, a letter written from a distant field operation where word processing facilities might reasonably be thought to be unavailable or an emergency appeal meant to convey haste and urgency. But in the majority of cases, where all other things are essentially equal, I'm no longer so sure that it makes a lot of difference.

money and get a higher response if you edited yours down to two pages? Typically, fundraisers test shortened letters hoping to save a few cents in a package's cost, and the result is often only a marginal net difference. But what would happen if you tried an eight-page or even twelve-page letter—assuming your message is sufficiently complex and compelling? Are there different inserts you could use, such as newspaper clippings, lift letters, or buckslips, that might raise the size of the average gift—or could you test *eliminating* such inserts to determine whether the savings outweigh any lowering in response?

For most nonprofits, each of these questions holds some promise of improving response significantly enough to justify the cost of testing it. There may well be half a dozen similarly attractive questions worth testing by any fundraising mailer.

Don't Skip the Basics

Although the following testing rules may seem elementary and obvious, forgetting them can throw a monkey wrench into your testing, thus sabotaging your results. Always remember to

- *Keep list coding easy*, and keep a written key to the coding system so test history can be tracked over the years and throughout personnel changes. Try to assign each list the same code in all mailings; this will make historical tracking simpler.
- *Keep tests clean*. Test only one thing at a time. Overzealous direct mail fundraisers can unwittingly invalidate their own tests by making “just one little change” to a package.

- *Keep the number of test cells manageable.* It's easy to get "testing fever" and go overboard with the number of test cells. Smart mailers keep it simple, using such models as the six-cell test: one control cell, four test cells that are variations on the control, and one new package cell—for example, a new format test. (In a format test, the ask, offer, and copy should be the same as they are in the control to make it a valid test, even when the reformatted package has a totally new design.)
- *Keep mailing your control even when testing.* Sometimes fundraisers learn about a new concept that's getting a lot of mileage for others and decide to roll out a similar package instead of their control before testing that new package in their market. But even though the concept works for other markets and prospects, it might not work for yours. Test, test, test. And *back-test* (that is, retest using a small test panel) your old controls; you may be pleasantly surprised.
- *Eliminate unusually large gifts* when measuring the revenue impact of a test. If most gifts from a mailing were in the neighborhood of \$25, a few random contributions of \$100 or more could easily skew the results. If the average gift was \$150, drop out any donations of \$1,000 or more.
- *Test even the best of packages.* A direct mail fundraising manager may be reluctant to make any changes in a package crafted by a top professional. But there's no reason not to do everything you can to turn it into an even bigger winner. It can really pay off to alter packages written by the "best in the business," adding elements or copy emphasis that have worked for other proven packages. Sometimes, "new and improved" wins the day. After all, that's what testing is all about!

Part II

Real-World Results

In the following six chapters I relate the results of tests in five areas that organizations often consider testing but that I have not yet discussed in any detail. Some of these tests were conducted by my agency and its clients, and some were carried out by colleagues at other organizations. These examples will help you get a better handle on what is worthwhile to test and give you a few hints about what you might expect when you conduct your own tests.

Signatures

Take a guess: if the text and artwork of two fundraising letters are otherwise identical in every respect, what difference in response would you expect if you changed the identity of the letter-signer on one of them?

Five percent? Twenty-five? Fifty?

For one leading consumer advocacy group, heavily dependent on direct mail fundraising revenue, this guessing game was not academic. The organization's founder and long-time executive director had just retired, after many years of high-profile coverage in the news media. Her replacement, though exceptionally able and articulate, was far less mediagenic and relatively unknown to the public. The organization's board members—and its direct mail fundraising counsel—were extremely nervous about changing anything so

basic as the identity of the letter-signer in the organization’s highly successful membership acquisition program.

The solution? A *test* to determine the extent of the damage if the organization made the change—and to buy time to find a celebrity who would sign the acquisition letter in the founder’s place if the damage were as great as the board feared.

Here’s how the test was structured: out of 172,000 individuals slated to receive membership acquisition letters, 60,000 were selected at random and divided into two statistically identical groups of 30,000 names each. The acquisition letters mailed to one group were signed by the founder, and the letters mailed to the other group—though identical in every other respect—were signed by the new executive director.

The results (Table 14.1) held a big surprise. *There was virtually no difference in the responses to the two versions of the letter!* The 30,000 letters mailed over the founder’s signature yielded 761 gifts. The 30,000 letters signed by the executive director generated a slightly lower response rate: 735 gifts. But the average of those contributions was a few cents higher than the average from the founder’s letters.

On balance the results for the two groups were statistically indistinguishable: \$388 per 1,000 letters mailed over the execu-

TABLE 14.1. Testing the Impact of a New Letter-Signer

Pkg.	Total Gifts	% Resp.	Avg. Gift	Rev./ 1,000
Old signer	761	2.54%	\$15.63	\$396
New signer	735	2.45	15.92	388
Difference with old signer	+26	+0.09	−0.29	+8
% difference with old signer	+4%	+4%	−2%	+2%

tive director's signature and \$396 per 1,000 for those signed by the founder.

Now, be careful about reading into this test more than the results support. First, no test results are 100 percent reliable. On a matter of this gravity, a nonprofit organization may be well advised to test at least *twice*. Second, the results of this test apply to *one* organization and, in fact, to only *one* direct mail fundraising letter. Response might have varied enormously if the two versions of the letter had differed in more ways than just the signer's identity.

But if you ever have to consider the impact of changing the letter-signer for *your* organization, you might keep the results of this test in mind. They're consistent with every similar test I've observed during my years of fundraising: generally speaking, the question of who signs the letter doesn't seem to mean a whole lot to donors.

In this respect, as in so many others, direct mail fundraising is counterintuitive. And the only way to *know* how people will really respond to your appeals is to test.

Teasers

Fads and rumors affect the thinking of direct mail fundraisers no less than that of financiers and politicians. One day simulated overnight express packages or 9-by-12-inch envelopes are “in,” and the next day they’re “out”—as often as not because “people say they’re not working anymore.”

Well, wise direct mail fundraisers actually *test* such suppositions. They know that test results will often challenge the fads, rumors, and conventional wisdom that may subtly or not so subtly sway decision making in our field.

Here, for example, is what happened when two organizations recently tested the notion that “teasers just aren’t working anymore.” (That’s one I’ve been hearing since I dropped my first mailing in 1979!)

One, an international human rights organization that I'll call Organization A, mailed two versions of an acquisition package. The versions were identical in every respect except for the artwork on the carrier envelopes. The test version bore the organization's logo and address in the *corner card* (the upper-left-hand corner), along with the typewritten name of its executive director. The carrier for the control version of the package featured reproductions of front-page stories about the release from prison of Nelson Mandela and Václav Havel, together with the bright red teaser, "You know their stories. But what about the others?" Each version was mailed to a test panel of approximately 16,500 names, 33,000 names altogether.

Not long after this mailing dropped, a local chapter of a national legal rights organization that I'll call Organization B conducted a similar test. The control package bore a boxed teaser to the right of the window: "IMPORTANT NOTICE: The most private decision you may ever have to make is going to be made for you . . . unless you act today." Both the control and the test versions featured the organization's logo and return address in the corner card, and both were printed on elegant blue granite stock. Exactly 25,000 were mailed of each version.

Months later, the jury came in: for these two organizations teasers work—or they bomb.

As you can see in Table 15.1, the test for Organization B yielded equivocal results: \$183 per 1,000 teaser packages mailed versus \$173 per 1,000 no-teaser packages. Dropping the teaser reduced revenue by nearly 6 percent. Certainly, that 6 percent could make a difference—but it's not enough more than the random variation in statistical results to be trustworthy. In other words, for this organization, it made sense to go on mailing the control; dropping the teaser is highly *unlikely* to improve its results.

TABLE 15.1. Testing Teaser Versus No-Teaser Carriers for Two Organizations

Pkg.	Total Gifts	% Resp.	Avg. Gift	Rev./ 1,000*
Org. A, teaser	75	0.46%	\$26.16	\$117
Org. A, no teaser	122	0.75	27.19	192
% difference with no teaser	+63%	+63%	+4%	+64%
Org. B, teaser	193	0.77	23.72	183
Org. B, no teaser	188	0.75	23.03	173
% difference with no teaser	−3%	−3%	−3%	−6%

In contrast, the results of Organization A's mailing were dramatic. The teaser package generated just \$117 per 1,000 (counting only gifts of less than \$100), whereas the no-teaser test version produced \$192 per 1,000 (plus a \$1,000 gift that was statistically meaningless but that nevertheless helped boost total revenue per 1,000 to \$270). Thus, for this organization, changing the carrier envelope artwork boosted revenue by 64 percent!

With such contrasting results, can we draw any lessons at all from these two tests? There is only one lesson that makes any sense to me: don't heed fads and rumors, because the results *you* obtain in the mail using similar techniques may be very different from someone else's.

Brochures



Quick: what's the *best* reason to mail a brochure along with your next direct mail fundraising letter?

1. You have a drawerful of nice photographs and no other way to use them.
2. The chair of your board of trustees insists on it.
3. You read somewhere that you're *supposed* to use brochures in direct mail.
4. You want to repeat and illustrate the benefits of sending a gift to your organization.

If you answered number 4, go to the head of the class. You may wear an amused and superior smile, if you wish. If you answered

numbers 1, 2, or 3—or if you want to find out how number 4 could be the wrong answer too—read on.

For starters, let's get something straight: under no circumstances I can imagine does it make sense to use a brochure *instead* of a fundraising letter. In a written appeal a letter is just about the only way to establish a personal connection between your organization and your donors. And if the letter does its job well, adding a brochure will usually have relatively little impact one way or the other. But there may be circumstances in which adding a brochure could lift results.

Consider, for example, the recent experience of two East Coast nonprofits that tested the value of adding brochures to direct mail acquisition packages.

For a nonprofit that focused on environmental concerns, Organization A, the inclusion of a brochure with dramatic photographs in an already impressive high-dollar prospecting package had an impact that could only be described as profound (see Table 16.1): response was 37 percent higher, and the average gift was 45 percent greater. Altogether, the brochure package yielded 98 percent more revenue than the no-brochure package!

As you can also see in Table 16.1, the experience of the second nonprofit, Organization B, was very different. A brochure illustrating the worldwide reach of the organization's field staff proved to have no meaningful impact on the results of a membership acquisition test mailing. Although the response was 7 percent higher with the brochure, the average contribution was 8 percent lower, canceling out the difference.

Why did these two organizations have such different experiences? And what might you be able to learn from them? Here's a plateful of food for thought:

TABLE 16.1. Testing Brochure Versus No-Brochure Packages for Two Public Interest Organizations

Pkg.	% Resp.	Avg. Gift	Rev./1,000
Org. A, brochure	2.65%	\$167.69	\$4,449
Org. A, no brochure	1.94	115.74	2,243
% difference with brochure	+37%	+45%	+98%
Org. B, brochure	0.48	27.21	130
Org. B, no brochure	0.45	29.39	132
% difference with brochure	+7%	−8%	−2%

- There's nothing inherently good or bad about using brochures in direct mail fundraising packages. If a brochure will strengthen your message—rather than distract your donors—it may be a good idea. But only if it actually *works*, by increasing response or raising the average gift, or both, is it definitely good.
- Every element in a direct mail fundraising package must be cost effective. So, your organization has to be able to afford the added cost of creating and producing a brochure. And that added cost has to be justified by increased revenue.

What do you do then if the chair of your board insists, regardless of the facts, that you include a brochure in that next direct mail fund appeal? I'll tell you what I'd do: I'd test it, even knowing that its impact may be negligible. Whoever said reality has anything to do with logic?

Window Envelopes

Doesn't everyone *know* that window envelopes are a dead giveaway for junk mail? Isn't it *obvious* that you can't expect good response from your donors if you send your appeals in window envelopes?

One skeptical fundraising consultant was unconvinced that anything in direct mail fundraising could be so obvious. This consultant (I'll call him "Felix") decided to find answers to these questions—by testing. And what he learned may be of value to anyone who solicits voluntary contributions by mail.

Felix persuaded a client, a candidate for the U.S. Senate, to *split-test* a recent appeal to previous donors. About half those in the target group were sent appeals in *closed-faced* outer envelopes (OEs), laser addressed on the front to match the identical addressing on both the letter and the reply device inside (in other words,

a *three-way match*). The remaining donors were sent otherwise identical packages but in *window envelopes*, through which the address on the reply device showed (a *two-way match*).

Though the number of packages the client mailed was small—fewer than 7,000 (3,229 to the first group and 3,250 to the second)—the number of responses received was comfortably large enough to be statistically significant, a total of 424 gifts after fifty-six days. The results can be seen in Table 17.1.

There was no difference in response for the two packages. But the window envelope version yielded gifts more than \$1 *higher* on average—and at a cost per piece that was more than \$0.06 *lower*. The result was a net advantage for the window envelope of \$108 per thousand—or nearly \$0.11 per package.

Although this advantage was not statistically meaningful, the fact that there was no demonstrated advantage for the closed-face envelope *was* important. It meant that spending more money for a three-way instead of a two-way match was simply a waste of money. However, the cost differential was not as great as Felix had

**TABLE 17.1. Testing Window
Versus Closed-Face Envelopes**

Pkg.	Total Gifts	% Resp.	Avg. Gift	Gross Rev./ 1,000	Net Rev./ 1,000
Closed-face OE, 3-way match	212	6.57%	\$62.82	\$4,124	\$3,247
Window OE, 2-way match	212	6.52	63.85	4,165	3,355
Difference with closed-face OE	0	+0.05	−1.03	−41	−108
% difference with closed-face OE	0%	+0.77%	−1.6%	−0.99%	−3%

anticipated. The difference in cost between the two-way match and the three-way match turned out to be a lot less than the difference in cost between a package with a two-way match and one requiring no matching. (Does this suggest a follow-up test?) However, Felix did learn from these results that at least in this case, reducing personalization did not reduce impact.

So, why did the “obvious” truth prove untrue? Felix’s theory was that these donors had given to the Senator’s campaign before and were willing to give again. They opened the envelope with the window that marked it as direct mail knowing that it was a solicitation to which they were likely to respond. That is, the donor already had some degree of commitment, and the added personalization of a closed-face (or the depersonalization of a window) was only a minor factor. In such situations the real test comes, Felix speculated, once donors are reading the inside materials. That is the point at which increased personalization might have a more noticeable impact.

Instinct tells me this theorizing is right on both counts—but I would add this cautionary note: to generalize about the applicability of direct mail testing can be costly. What worked for this client may not work for your nonprofit organization—or even for another political candidate. And what worked well in one appeal for Felix’s client might not work so well in another appeal for the same client.

Are test results worth nothing then? Not on your life! Because the next time Felix conducted a personalized donor appeal, he thought twice before spending the extra money for a three-way match! Perhaps you should too.

My agency’s experience supports this caution. In two head-to-head tests conducted on two subsequent control packages two years apart, one client tested the proposition that closed-face

envelopes will generate better responses than window envelopes. The remarkably similar results from mailings to panels of 30,000 each and then 40,000 each revealed a somewhat higher response rate and greater revenue for the closed-face package. However, the marginal advantage gained fell significantly short of convincing statistical proof. Given the extra cost of the closed-face envelope, the margin did not seem great enough to justify the added financial exposure.

For an acquisition mailing for another client, involving two test panels of 40,000 names each, the closed-face carrier yielded twenty more gifts, a lift of 10 percent that nevertheless was not statistically significant. Moreover, although the closed-face envelope brought in \$6,680 and the window envelope \$5,171, a difference of \$1,509, the closed-face envelope cost \$29 per thousand more (in this case, adding \$1,160 to the costs). Finally, a full \$1,300 of the \$6,680 from the closed-face envelope came in the form of gifts of \$100 or more—which ought not to be counted for the purpose of statistical evaluation. (As discussed in the perspective for Chapter Eight, when most of the gifts are small, the larger gifts are statistically random and cannot be relied on to recur.) The window envelope package yielded only \$700 in gifts of \$100 or more. When the \$100 gifts were factored out, the client could attribute only \$4,471 to the window and \$5,380 to the closed-face envelope. Because the latter cost \$1,160 more, the true revenue comparison showed that in this instance the window envelope actually produced *better* results.

Keep testing, though! Who knows what the next try will show?

Mailing for the Environment (I)

After Earth Day 1990 and the flood of environmental books and articles it occasioned, you might have thought all the direct mail recipients on this planet would be ready to form a lynch mob and look for the nearest fundraiser. But apart from the angry public comments of an outspoken few during this period of heightened sensitivity to the waste of paper, what *did* donors think about the amount of paper used in direct mail packages?

For years many nonprofit organizations have had anecdotal evidence—in the form of nasty calls or notes from members—that some direct mail donors are incensed about such things as duplicate letters and liberal use of white space in package layouts. But a nonprofit organization based in the Midwest decided to test the effect of one such issue on donors' actual *behavior* not long after

Earth Day 1990. The lesson that nonprofit learned may be instructive for you too.

This organization had been mailing an acquisition package nationwide for more than two years—with remarkable results. Although the response rate was usually less than three-quarters of 1 percent, charter membership contributions were averaging about \$50, double the initial gifts typically received by similar organizations. As a result, every prospect mailing undertaken by the group had posted a profit, and its file of major donors (defined as \$100-and-up givers) was growing at a rapid rate.

It wasn't at all difficult to understand why the campaign was working so well. The organization's program filled an obvious and important need. The charter membership offer featured uniquely attractive benefits, both tangible and intangible. Prospect lists were carefully chosen through intensive research. And the package's upscale copy and design were both compelling.

But no self-respecting direct mail fundraiser is content with success. In a continuing effort to improve these already impressive results, the organization decided to test the effect of reducing the amount of paper in its acquisition packages. In the control the four-page letter occupied four separate sheets of paper, each sheet printed on one side only. The organization opted for a split-test to determine whether printing the very same text back-to-back on two sheets would improve the net results. And this was not a half-hearted test: it involved a total of 210,000 packages, 105,000 of each of the two formats.

The organization's reasoning couldn't have been sounder: first, in rollout quantities, printing only two sheets of paper would lower the package cost by \$21 per thousand (about 6 percent), and second, the public's growing environmental consciousness about paper wasted in direct mail might increase response.

Well, that's not quite how things worked out. The four-sheet control yielded \$33,928 in revenue. The two-sheet test produced a similar average gift but a somewhat lower response and so yielded only \$31,260—nearly \$2,700 less than the control (see Table 18.1).

Given these numbers, it might seem that the four-sheet version won the test. But the difference in the number of gifts generated by the two panels was not convincingly large. With a confidence level of only 86 percent, you might just as well expect the ratio to reverse the second time around. In other words, *there was no discernable difference in response*.

But that's not all. The printing and lettershop savings from eliminating two sheets of paper from every 105,000 packages in a rollout mailing would have been about \$2,205—not as much as the actual difference in revenue between the two packages, but a meaningful amount nonetheless, and one with the effect of boosting the net revenue from the two-sheet letter.

In these circumstances, despite the huge size of the test panels in this mailing—I would advise retesting. If it's true that the two-sheet version actually generates fewer gifts, that roughly \$20 per thousand savings in cost could look awfully costly!

TABLE 18.1. Testing 4-Sheet Versus 2-Sheet Letters with Identical Text

Pkg.	Total Gifts	% Resp.	Avg. Gift	Gross Rev.
Control: 4-sheet letter	662	0.63%	\$51.25	\$33,928
Test: 2-sheet letter	609	0.58	51.33	31,260
Difference with 4-sheet letter	+53	+0.05	−0.08	+2,668
% difference with 4-sheet letter	+9%	+9%	−0.2%	+9%

There's no hard-and-fast conclusion to be drawn from these test results. But it's fair to conclude that acting in an environmentally conscious manner won't *necessarily* gain you the undying love and devotion of direct mail donors and thus improve your results. As always, it makes sense to *test* before making even the most logical and attractive changes in a successful direct mail package.

Mailing for the Environment (II)

In the preceding chapter I reported on one Midwestern non-profit's unsuccessful effort to capitalize on the American public's growing environmental awareness. By reducing the number of sheets of paper in its direct mail acquisition packages, it had hoped to gain donors' respect and also lower production costs. But what this fundraiser expected was not what it got.

At about the same time, a West Coast environmental organization was undertaking an interesting test of a related question: Will direct mail donors respond more generously if they're forcefully told that the paper in an acquisition package is recycled?

The outer envelope (OE) for the organization's control package included the following legend in small but clearly legible green type beneath the window: "RECYCLED PAPER—WINDOW IS RECYCLABLE GLASSINE." To learn whether this legend had the positive effect

the organization assumed it did, the organization *omitted* it from a test version of the package. As part of a larger acquisition mailing, some 60,000 names were selected for a test pool, which was divided at random into two panels of approximately 30,000 names each. All 60,000 packages mailed to these panels were identical in every respect except that the OE “RECYCLED PAPER” legend was printed on those mailed to one group and omitted from those mailed to the other.

Please note that this was *not* a test of the effect of recycled paper itself. Both the control and the test package were printed entirely on recycled stock. The phrase “RECYCLED PAPER” appeared in inconspicuous type in three places *inside* each of the two packages.

What was the impact of this subtle effort to satisfy the environmental sensitivity of direct mail donors specially selected for their likely interest in environmental issues? There was *no* significant difference in results between the two groups.

The control group, which received the “RECYCLED PAPER” message on the envelope, sent 159 gifts, totaling \$4,271. The test group contributed 162 gifts, totaling \$4,261. By every measure the difference in results was statistically meaningless (see Table 19.1).

Without intensive market research, such as focus groups or a telephone survey of respondents, we can only speculate why this environmentally sensitive population was entirely unaffected by a gentle reminder that an environmental group was willing to put its money where its mouth was. There are at least four possible explanations:

1. Perhaps the “RECYCLED PAPER” legend was simply unnecessary, because environmental donors *assume* that an environmental organization would use only recycled

TABLE 19.1. Testing the Value of an Environmental Message on Environmental Donors

Pkg.	Total Gifts	% Resp.	Avg. Gift	Gross Rev.
Test, omit OE message	162	0.54%	\$26.30	\$4,261
Control, include OE message	159	0.53	26.86	4,271
Difference without OE message	+3	+0.01	−0.56	−10
% difference	+2%	+2%	−2%	−2%

paper in its direct mail program. (This is in fact true of this organization.)

2. Perhaps the repeated use of the shorter phrase, “RECYCLED PAPER,” inside the package was sufficient.
3. Perhaps the message was entirely too subtle, the type too small and unobtrusive.
4. Or *perhaps* only a very small and vocal minority (including staff members of direct mail fundraising agencies!) really care about this issue—those who would take the trouble to write and complain or to bend the executive director’s ear after a meeting or at a party.

In the absence of a lot more data, it’s impossible to know why the test turned out as it did. But there is at least one more general conclusion that may be drawn from this test: no matter how smart we are, we can’t assume we know how direct mail donors will react to any of our tactics and devices.

Part III

Testing in Action

In the foregoing two parts of this book, I've included examples of actual test results wherever possible. But this is the part where I've pulled out all the stops. I've plundered my agency's "Testing Bible" to relate the results of 286 separate direct mail fundraising tests representing my and my colleagues' work for thirty-one clients. I'm certain this part alone contains more actual fundraising test results than have ever before appeared in print in one place.

Each of the tables in Part Three contains information about a series of tests. Each set of results in each series compares the mailings, usually labeled A and B, sent to two test panels. In instances where different packages were sent to more than two panels, the most telling results may be compared (say, for Packages A and C or Packages B and D) or the results may be given as a series of comparisons (say, Package A compared to B and then Package A compared to C). You will see a number of these series in Table 26.1, “Testing the Ask,” in particular.

The first row in each test series addresses the purpose of the mailing (for example, donor acquisition or member renewal). The second and third rows describe the difference between the two packages mailed (for example, a newspaper insert versus no insert). The three columns to the right of this information show the quantity of each package mailed (Qty.), the cost per one thousand packages mailed (Cost/1,000), and the total number of gifts received (Total Gifts). The next two columns address the test of statistical significance described in Chapter Eight. They show, first, the numerical difference between the total gifts received from each of the two panels (Diff. Btwn. Gifts) and, second, the square root of the sum of the total gifts for both panels multiplied by 2 (Sq. Root of Total Gifts $\times$ 2). When the first figure is larger than the second, the test results are likely to be statistically significant.

The last six columns on the right show the total revenue (Total Rev.), the average gift (Avg. Gift), the total cost (Total Cost), the net revenue (Net Rev.), the average cost of acquiring one donor (Acq. Cost/Donor), and the percentage of people who responded (% Resp.). A negative number in the acquisition cost column indicates that the net revenue per donor was less than the cost of acquisition. A positive number indicates a profit on the mailing.

Below each set of results, you will find the confidence level (discussed in Chapter Nine) for that test and the tentative conclusion drawn *from that test for that client at that time*.

My hope is that this treasure trove of testing data will help you make up your own mind about what's worth testing and what isn't. Here, you'll find some patterns of results—and at times a clear *lack* of patterns. Here too you should find a number of ideas worth testing for your own organization.

But I hope you will *not* treat these examples as gospel, as my colleagues and I so irreverently refer to them. You will do yourself and your organization a great disservice if you regard the results revealed here as predictions of the results your organization will obtain.

Keep in mind that no test yields results that are valid for more than one organization under one set of controlled circumstances at one point in time. When you extrapolate far beyond those parameters, you run a big risk.

There are no shortcuts to success in direct mail fundraising. Testing is not a source of immutable truth—nevertheless it is one of the essential tools in the fundraiser's toolbox.

Inserts

The conventional wisdom in direct mail—whether it is intended to raise funds or has almost any other purpose—is to include a brochure or some other form of insert in the acquisition package. As the logic has it, an insert is needed to supply additional details about the offer described in the letter; to illustrate the offer's benefits, often graphically; to highlight some special incentive, often a back-end premium, to respond to the appeal; to present an opportunity for prospects to become involved in some way other than merely by sending a check; or simply to support the arguments raised in the letter, sometimes by featuring a celebrity endorsement.

Does this conventional wisdom hold water? Table 20.1 presents the results of twenty-five donor acquisition tests for twelve of my agency's clients in recent years. Overall these tests suggest

that this conventional wisdom is rather leaky. Nevertheless, a few organizations experienced dramatic boosts in response when they used inserts. Should you test an insert for your organization's mailings? Judge for yourself as you review these results.

You'll note that the inserts tested are not normally described by type of insert. The omission of descriptive detail is deliberate. There are no useful conclusions you can draw from a test that compares two specific inserts against each other. But it is valuable to understand the extent to which changes of the sort shown here can have a significant impact on results.

TABLE 20.1. Testing Inserts

		Sq. Root of Total				Acq. Cost/ Donor			
	Qty.	Cost/ 1,000	Total Gifts	Diff. Btwn. Gifts	Total Gifts × 2	Total Rev.	Avg. Gift	Total Cost	Net Rev.
Client 1 Donor Acquisition Mailing									
A Newspaper insert	20,056	\$497	176			\$7,888	\$45	\$9,968	-\$2,080
B No insert	20,050	471	139	37	35	4,891	35	9,444	-4,553
Confidence level: 96%; statistically significant.									
Tentative conclusion: newspaper insert lifts response.									
Client 2 Donor Acquisition Mailing									
A Decal	50,000	419	497			23,190	47	20,950	2,240
D No decal	50,000	419	462	35	62	15,004	32	20,950	-5,946
Confidence level: less than 80%; you might as well flip a coin.									
Tentative conclusion: no statistically significant difference; decal may be an unnecessary additional expense, but retest because of higher average gift.									
Client 3 Donor Acquisition Mailing									
A No sticker	39,974	355	309			11,228	36	14,191	-2,963
B Sticker	39,974	412	401	92	53	14,769	37	16,469	-1,700
Confidence level: 99%; it doesn't get much better than this!									
Tentative conclusion: sticker lifts response.									

Client 5 Donor Acquisition Mailing

A Insert	30,000	512	163		5,727	35	15,360	-9,633	-59	0.543
B No insert	30,000	462	168	5	36	7,348	44	13,860	-39	0.560

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference; insert is an unnecessary additional expense.

Client 1 Donor Acquisition Mailing

A No insert	22,000	485	117		4,653	40	10,670	-6,017	-51	0.532
B Insert	22,000	525	118	1	31	5,362	45	11,550	-52	0.536

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference; but retest because of higher average gift.

Client 5 Donor Acquisition Mailing

A No insert	33,036	474	213		9,854	46	15,659	-5,805	-27	0.645
B Photo insert	33,027	506	243	30	43	10,930	45	16,712	-24	0.736

Confidence level: 84%; unreliable, but may be worth retesting.

Tentative conclusion: no statistically significant difference, but retest photo insert.

Client 2 Donor Acquisition Mailing

B Insert	25,000	350	291		7,930	27	8,750	-820	-3	1.164
C No insert	25,000	340	274	17	48	7,180	26	8,500	-5	1.096

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference; insert is an unnecessary additional expense.

TABLE 20.1. Testing Inserts, Continued

Sq. Root of Total Gifts									
Qty.	Cost/ 1,000	Total Gifts	Diff. Btwn. Gifts	Total Gifts × 2	Avg. Gift	Total Cost	Net Rev.	Acq. Cost/ Donor	% Resp.
Client 6 Donor Acquisition Mailing									
C Anniv. stamps insert	40,000	\$383	266		\$26	\$15,320	-\$8,500	-\$32	0.665%
D No insert	40,000	358	278	47	30	14,320	-6,001	-22	0.695
Confidence level: less than 80%; you might as well flip a coin.									
Tentative conclusion: no statistically significant difference; stamps are an unnecessary additional expense.									
Client 7 Donor Acquisition Mailing									
A No insert (control)	30,000	313	256		29	9,390	-1,955	-8	0.853
B Photo insert	30,000	313	238	44	30	9,390	-2,261	-10	0.793
Confidence level: less than 80%; you might as well flip a coin.									
Tentative conclusion: no statistically significant difference; insert is an unnecessary additional expense.									
Next entry continues this test.									
Client 7 Donor Acquisition Mailing									
A No insert (control)	30,000	313	256		29	9,390	-1,955	-8	0.853
C In memory/honor insert	30,000	313	222	44	27	9,390	-3,393	-15	0.740
Confidence level: 87%; unreliable, but probably worth retesting.									
Tentative conclusion: insert is an unnecessary additional expense.									

Client 8 Donor Acquisition Mailing

A No insert	17,001	399	106		2,796	26	6,783	-3,987	-38	0.623
B Insert	16,999	412	114	8	30	2,637	7,004	-4,367	-38	0.671

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference; insert is an unnecessary additional expense.

Client 4 Donor Acquisition Mailing

A No insert	24,988	589	590		19,126	32	14,718	4,408	7	2.361
B Insert	24,987	583	558	32	68	16,881	14,567	2,314	4	2.233

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference; insert is an unnecessary additional expense.

Client 9 Donor Acquisition Mailing

A Name labels (control)	30,000	394	192		4,545	24	11,820	-7,275	-38	0.640
B Name labels with insert	30,000	404	184	8	39	4,669	12,120	-7,451	-40	0.613

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference; insert is an unnecessary additional expense.

Client 7 Donor Acquisition Mailing

A No insert (control)	30,000	374	236		6,343	27	11,220	-4,877	-21	0.787
B Insert	30,000	374	219	17	43	6,192	11,220	-5,028	-23	0.730

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference; insert is an unnecessary additional expense.

Client 9 Donor Acquisition Mailing

A	Info insert (control)	50,006	380	449		13,321	30	19,002	-5,681	-13	0.898	
C	Photo insert (control)	49,988	380	408	41	59	10,429	26	18,995	-8,566	-21	0.816

Confidence Level: 83%; unreliable, but may be worth retesting.

Tentative conclusion: no statistically significant difference, but retest.

Client 12 Donor Acquisition Mailing

A	Insert	24,988	474	110		2,253	20	11,838	-9,585	-87	0.440	
B	No insert	24,986	474	115	5	30	2,028	18	11,837	-9,810	-85	0.460

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference; insert is an unnecessary additional expense.

Client 2 Donor Acquisition Mailing

D Inserts	29,596	435	227		6,538	29	12,874	-6,336	-28	0.767	
E No inserts	29,595	435	214	13	42	5,390	25	12,874	-7,484	-35	0.723

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference; inserts are an unnecessary additional expense.

Client 2 Donor Acquisition Mailing

A	No decal	30,000	379	254		7,162	28	11,370	-4,208	-17	0.847	
C	Decal	30,000	409	282	28	46	7,280	26	12,270	-4,990	-18	0.940

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference; decal doesn't increase response, but retest.

TABLE 20.1. Testing Inserts, Continued

Sq. Root of										
Qty.	Cost/ 1,000	Total Gifts	Diff. Btwn. Gifts	Total		Avg. Gift	Total Cost	Net Rev.	Acq. Cost/ Donor	% Resp.
				Total Gifts	× 2					
Next entry continues this test.										
Client 2 Donor Acquisition Mailing										
A	No decal	30,000	\$379	254		\$28	\$11,370	-\$4,208	-\$17	0.847%
D	Decal (control)	30,000	379	252	2	30	11,370	-3,743	-15	0.840
Confidence level: less than 80%; you might as well flip a coin.										
Tentative conclusion: no statistically significant difference.										

Postage

Are you one of those people who automatically tosses envelopes into the trash if they don't bear first-class postage? Don't laugh: that behavior has become more common than you might think. If it is true for you too, you might have a better visceral understanding than other people do of one reason why the amount or type of postage on a fundraising appeal could have an impact on the results.

There are other, perhaps even better, reasons, of course. First-class mail is delivered more reliably than third-class—and is forwarded by the postal service rather than unceremoniously dumped if it's undeliverable. In any case it's natural to assume that all other things being equal, an appeal sent via first-class mail will yield more revenue than one sent via nonprofit bulk mail. But (here's the rub) *how much* more revenue? Enough to more than compensate for

the substantial additional expense of using first-class rather than third-class postage?

That is one of the questions my colleagues and I have explored in direct mail tests with many of our clients over the years. In Table 21.1, you'll see the results of eleven such tests, examining whether first-class postage sufficiently outpulls nonprofit third-class, gauging the difference between metered and stamped postage, and attempting to answer other questions involving postage.

Our findings in these tests suggested that the form of postage does not *usually* make a difference but that this is by no means a hard-and-fast rule. Draw your own conclusions from the numbers that follow!

TABLE 21.1. Testing Postage

		Sq. Root of				Total				Acq.			
		Qty.	Cost/ 1,000	Total Gifts	Diff. Btwn. Gifts	Total Gifts × 2	Total Rev.	Avg. Gift	Total Cost	Net Rev.	Donor	% Resp.	
Client 1 Donor Acquisition Mailing													
A	Nonprofit stamp	35,094	\$415	102			\$6,611	\$65	\$14,564	-\$7,953	-\$77.97	0.291%	
B	Nonprofit meter	35,087	415	101	1	28	3,502	35	14,561	-11,059	-109.50	0.288	
Confidence level: less than 80%; you might as well flip a coin.													
Tentative conclusion: no statistically significant difference; nonprofit stamp is an unnecessary additional expense.													
Client 3 Donor Acquisition Mailing													
A	Nonprofit stamp	29,043	391	242			8,601	36	11,356	-2,755	-11.38	0.833	
B	Nonprofit meter	29,035	384	242	0	44	8,573	35	11,149	-2,576	-10.65	0.833	
Confidence level: less than 80%; you might as well flip a coin.													
Tentative conclusion: no statistically significant difference; nonprofit stamp is an unnecessary additional expense.													
Client 13 Donor Acquisition Mailing													
A	Nonprofit stamp	20,891	380	106			5,931	56	7,939	-2,008	-18.94	0.507	
B	Nonprofit meter	20,889	380	108	2	29	8,437	78	7,938	499	4.62	0.517	
Confidence level: less than 80%; you might as well flip a coin.													
Tentative conclusion: no statistically significant difference; stamp is an unnecessary additional expense.													

Client 4 Donor Acquisition Mailing

A Nonprofit postage	25,000	447	321		9,178	29	11,175	-1,997	-6.22	1.284
B 1st-class postage	25,000	584	377	56	53	10,493	28	14,600	-4,107	-10.89

Confidence level: 96%; statistically significant.

Tentative conclusion: first-class postage increases response.

Client 6 Donor Acquisition Mailing

E 1st-class postage	15,000	527	93		2,307	25	7,905	-5,598	-60	0.620
C Nonprofit postage	15,000	369	90	3	27	2,177	24	5,535	-3,358	-37

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference; first-class postage is an unnecessary additional expense.

Client 6 Donor Acquisition Mailing

A Nonprofit postage	30,002	385	53		1,286	24	11,551	-10,265	-194	0.176
B 1st-class postage	29,998	571	114	61	26	2,301	20	17,129	-14,828	-130

Confidence level: 99%; it doesn't get much better than this!

Tentative conclusion: first-class postage increases response.

Client 5 Donor Acquisition Mailing

A Nonprofit meter	30,004	370	218		6,383	29	11,116	-4,733	-22	0.727
B Nonprofit stamp	29,996	371	202	16	41	5,589	28	11,114	-5,526	-27

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference; nonprofit stamp is an unnecessary additional expense.

TABLE 21.1. Testing Postage, Continued

Sq. Root of Total												
Qty.	Cost/ 1,000	Total Gifts	Diff. Btwn. Gifts	Total Gifts × 2	Avg. Gift	Total Cost	Net Rev.	Acq. Cost/ Donor	% Resp.			
Client 7 Member Renewal Series												
A	1st-class postage	19,081	\$660	991		\$52,590	\$53	\$12,593	\$39,997	\$40	5.194%	
B	Nonprofit postage	19,081	520	1,027	36	90	55,498	54	9,922	45,575	44	5.382
Confidence level: less than 80%; you might as well flip a coin.												
Tentative conclusion: no statistically significant difference; first-class postage is an unnecessary expense.												

Outer Envelopes

Traditional practitioners of direct mail spend a lot—perhaps an inordinate amount—of their time on the appearance of the outer, or carrier, envelope (OE). In Table 22.1, you can see the results of twenty-three tests my colleagues and I conducted for a number of our clients in recent years in order to explore several aspects of this question.

These clients tested variations in the size and color of the outer envelope, the use of closed-face as opposed to window envelopes, and the presence or absence of a printed message on the envelope. Possible tests that are not included in this batch of examples might involve printing a message on the back of the envelope in addition to the one on the front, using photographs or other graphic images on the outside, and placing the letter-signer's signature or typed name in the corner card (upper-left-hand corner), among others.

Once again we found that changes in the outer envelope often made little difference in response, although there were exceptions to this pattern. As you may have noticed in the previous two tables in Part Three, the same organization may run two or more similar tests, even when the results have been unpromising, looking repeatedly for the element that will improve a package either incrementally or dramatically. You can see such patterns in this table too. Through such ongoing testing, an acquisition package can be refined over time to a higher level of performance.

TABLE 22.1. Testing Outer Envelopes

		Sq. Root of				Total				Acq.			
		Qty.	Cost/ 1,000	Total Gifts	Diff. Btwn. Gifts	Total Gifts × 2	Total Rev.	Avg. Gift	Total Cost	Net Rev.	Donor	% Resp.	
Client 7 Donor Acquisition Mailing													
A	White OE (control)	30,001	\$387	265			\$6,482	\$24	\$11,610	-\$5,128	-\$19	0.883%	
B	Canary OE	30,000	387	260	5	46	5,892	23	11,610	-5,718	-22	0.867	
Confidence level: less than 80%; you might as well flip a coin.													
Tentative conclusion: no statistically significant difference; canary envelope makes no difference.													
Client 7 Donor Acquisition Mailing													
A	White OE (control)	50,008	394	576			12,339	21	19,703	-7,364	-13	1.152	
B	Canary OE	50,000	394	545	31	67	11,749	22	19,700	-7,951	-15	1.090	
Confidence level: less than 80%; you might as well flip a coin.													
Tentative conclusion: no statistically significant difference; canary envelope makes no difference.													
Client 14 Donor Acquisition Mailing													
A	White OE (control)	20,000	491	212			4,917	23	9,820	-4,903	-23	1.060	
B	Canary OE	20,000	491	182	30	40	3,241	18	9,820	-6,579	-36	0.910	
Confidence level: 87%; unreliable, but probably worth retesting.													
Tentative conclusion: no statistically significant difference; canary envelope makes no difference.													

Tentative conclusion: no statistically significant difference; two-color envelope is an unnecessary additional expense.

Client 15 Donor Acquisition Mailing

B #10 blank OE	50,000	366	462		14,198	31	18,300	-4,102	-9	0.924	
C #10 OE with teaser and design	50,000	383	360	102	57	8,384	23	19,150	-10,766	-30	0.720

Confidence level: 99%; it doesn't get much better than this!

Tentative conclusion: blank #10 outer envelope increases response.

Client 15 Donor Acquisition Mailing

A Large OE (control)	40,000	381	247		6,451	26	15,240	-8,789	-36	0.618	
B #10 OE	40,000	365	268	21	45	7,713	29	14,600	-6,887	-26	0.670

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference; large outer envelope is an unnecessary additional expense.

Client 6 Donor Acquisition Mailing

A Window OE	40,000	418	213		5,371	25	16,720	-11,349	-53	0.533	
B Closed-face OE	40,000	447	234	21	42	6,955	30	17,880	-10,925	-47	0.585

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference; closed-face outer envelope is an unnecessary additional expense.

Client 16 Donor Acquisition Mailing

A White OE (control)	50,000	401	610		16,193	27	20,050	-3,857	-6	1.220	
B Brown kraft OE	49,999	411	554	56	68	14,340	26	20,550	-6,210	-11	1.108

Confidence level: 90%; probably significant, but you might want to retest.

Tentative conclusion: kraft envelope is an unnecessary additional expense.

Client 7 Donor Acquisition Mailing

A Printed OE (control)	30,000	266	421		11,746	28	7,990	3,756	9	1.403	
B Blank OE	30,000	266	387	34	57	11,586	30	7,990	3,596	9	1.290

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference; printing on an outer envelope is an unnecessary additional expense.

Client 6 Donor Acquisition Mailing

A White OE	30,000	380	156		4,196	27	11,400	-7,204	-46	0.520	
B Brown kraft OE	30,000	380	172	16	36	4,083	24	11,400	-7,317	-43	0.573

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference; brown kraft outer envelope is an unnecessary additional expense.

Client 5 Donor Acquisition Mailing

A Printed OE (control)	30,011	429	344		10,647	31	12,865	-2,218	-6	1.146	
B Blank OE	30,017	429	321	23	52	9,721	30	12,867	-3,146	-10	1.070

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference; printing on an outer envelope is an unnecessary additional expense

Client 11 Donor Acquisition Mailing

A #10 white OE	25,006	337	305		4,762	16	8,427	-3,665	-12	1.220	
B #10 brown kraft OE	25,002	338	293	12	49	4,638	16	8,451	-3,813	-13	1.172

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference; brown kraft outer envelope is an unnecessary additional expense.

TABLE 22.1. Testing Outer Envelopes, Continued

Sq. Root of Total										
Qty.	Cost/ 1,000	Total Gifts	Diff. Btwn. Gifts	Total Gifts × 2	Avg. Gift	Total Cost	Net Rev.	Acq. Cost/ Donor	% Resp.	
Client 11 Donor Acquisition Mailing										
C	6" x 9" white OE	24,998	335	336		\$5,103	\$15	\$8,874	−\$11	1,344%
D	6" x 9" brown kraft OE	24,994	355	337	1	52	16	8,873	−11	1,348
Confidence level: less than 80%; you might as well flip a coin. Tentative conclusion: no statistically significant difference.										
Client 11 Donor Acquisition Mailing										
A	#10 white OE	25,006	337	305		4,762	16	8,427	−12	1,220
C	6" x 9" white OE	24,998	355	336	31	51	15	8,874	−11	1,344
Confidence level: less than 80%; you might as well flip a coin. Tentative conclusion: no statistically significant difference; larger envelope is an unnecessary additional expense.										
Client 11 Donor Acquisition Mailing										
B	#10 brown kraft OE	25,002	338	293		4,638	16	8,451	−13	1,172
D	6" x 9" brown kraft OE	24,994	355	337	44	50	16	8,873	−11	1,348
Confidence level: 92%; probably significant, but you might want to retest. Tentative conclusion: larger envelope may increase response, but retest because of extra cost.										

Client 17 Donor Acquisition Mailing

A Printed OE (control)	20,007	453	82		2,056	25	9,065	-7,009	-85	0.410	
B Blank OE	19,993	453	109	27	28	2,195	20	9,059	-6,864	-63	0.545

Confidence level: 95%; statistically significant.

Tentative conclusion: plain outer envelope increases response.

Client 8 Donor Acquisition Mailing

A Unbleached OE	30,654	363	174		4,931	28	11,119	-6,189	-36	0.568	
B Brown bag OE	30,637	363	181	7	38	4,599	25	11,113	-6,514	-36	0.591

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference.

Client 6 Donor Acquisition Mailing

A Blue granite OE	30,000	377	239		5,761	24	11,313	-5,552	-23	0.796	
C White OE	30,000	377	232	7	43	6,158	27	11,313	-5,155	-22	0.774

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference; blue granite outer envelope is an unnecessary additional expense.

Outer Envelope Teasers

Since time immemorial, or 1965 (whichever came first), direct mail fundraising practitioners have labored under the impression that direct mail appeals must include some sort of provocative legend or statement on the outer envelope. Something intriguing or challenging or puzzling or greed provoking, or—whatever. You know the kind of messages I’m talking about, of course:

- FREE GIFT ENCLOSED
- Photos enclosed. Do not bend.
- Are you ready to take the next step?
- 2003 Annual Renewal Notice

The list is endless. And direct mail copywriters rack their brains for hours and even days on end to craft these little gems of wordsmithing.

But do they work? Or does one teaser do better than another?

Table 23.1 presents results from twenty-five teaser tests conducted by my agency in recent years. Outer envelope teasers have been known to pay off, although more often even the carefully crafted ones seem to have little effect on recipients. You'll have a better idea of the typical effectiveness of teasers after you've looked at these numbers—but of course you won't know how effective your own new teaser is until you test it.

TABLE 23.1. Testing Outer Envelope Teasers

		Sq. Root of Total			Total			Acq. Cost/ Donor		
	Qty.	Cost/ 1,000	Total Gifts	Diff. Btwn. Gifts	Total Gifts × 2	Total Rev.	Avg. Gift	Total Cost	Net Rev.	% Resp.
Client 15 Donor Acquisition Mailing										
A Teaser	124,473	\$361	947			\$30,740	\$32	\$44,935	-\$14,195	0.761%
B No teaser	124,446	361	860	87	85	29,084	34	44,925	-15,841	0.691
Confidence level: 95%; statistically significant.										
Tentative conclusion: teaser lifts response.										
Client 9 Donor Acquisition Mailing										
B Teaser	50,000	416	576			14,751	26	20,800	-6,049	1.152
C No teaser	50,000	416	583	7	68	13,281	23	20,800	-7,519	1.166
Confidence level: less than 80%; you might as well flip a coin.										
Tentative conclusion: no statistically significant difference; teaser does not increase response.										
Client 9 Donor Acquisition Mailing										
A No teaser	30,000	419	397			9,049	23	12,570	-3,521	1.323
C Teaser	30,000	419	346	51	55	9,583	28	12,570	-2,987	1.153
Confidence level: 94%; probably significant, but you might want to retest.										
Tentative conclusion: teaser does not increase response.										

Sq. Root of Total										
		Cost/	Total	Diff. Btwn.	Total	Avg.	Total	Net Rev.	Acq.	%
Qty.		1,000	Gifts	Gifts	Gifts × 2	Gift	Cost		Cost/	Resp.
Client 6 Donor Acquisition Mailing										
	C Teaser	30,000	\$379	189		\$22	\$11,370	-\$7,126	-\$38	0.630%
	D No teaser	30,000	379	176	13	32	11,370	-5,662	-32	0.587
Confidence level: less than 80%; you might as well flip a coin.										
Tentative conclusion: no statistically significant difference; teaser does not increase response.										
Client 18 Donor Acquisition Mailing										
	A No teaser	23,579	486	132		45	11,459	-5,551	-42	0.560
	B Teaser	23,567	486	93	39	38	11,454	-7,914	-85	0.395
Confidence level: 99%; it doesn't get much better than this!										
Tentative conclusion: OE with no teaser lifts response.										
Client 3 Donor Acquisition Mailing										
	A No teaser	39,949	390	342		44	15,580	-699	-2	0.856
	B Teaser	39,942	390	318	24	42	15,577	-2,296	-7	0.796
Confidence level: less than 80%; you might as well flip a coin.										
Tentative conclusion: no statistically significant difference; teaser does not increase response.										

Client 5 Donor Acquisition Mailing

A No teaser	29,978	495	213		11,747	55	14,839	-3,092	-15	0.711
B Teaser	29,970	495	158	55	39	8,124	14,835	-6,711	-42	0.527

Confidence level: 99%; it doesn't get much better than this!

Tentative conclusion: teaser lowers response.

Client 8 Donor Acquisition Mailing

A No teaser	31,492	400	201		7,343	37	12,597	-5,254	-26	0.638
B Teaser	31,483	400	193	8	40	6,593	12,593	-6,000	-31	0.613

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference; teaser does not increase response

Client 13 Donor Acquisition Mailing

A No teaser	18,190	355	73		3,450	47	6,457	-3,007	-41	0.401
B Teaser	18,182	355	92	19	26	4,672	6,455	-1,783	-19	0.506

Confidence level: 86%; unreliable, but probably worth retesting.

Tentative conclusion: teaser does not increase response, but retest this teaser.

Client 13 Donor Acquisition Mailing

A Teaser	18,847	412	78		3,390	43	7,765	-4,375	-56	0.414
B No teaser	18,836	412	91	13	26	4,960	7,760	-2,800	-31	0.483

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference; teaser does not increase response.

Client 19 Donor Acquisition Mailing

A Control teaser	30,000	345	212		6,513	31	10,350	-3,837	-18	0.707
B New teaser	30,000	345	209	3	41	30	10,350	-4,031	-19	0.697

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference; new teaser does not increase response.

Client 3 Donor Acquisition Mailing

A Teaser 1	20,000	400	132		4,788	36	8,000	-3,212	-24	0.660
B Teaser 2	20,000	400	146	14	33	41	8,000	-2,047	-14	0.730

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference; test without teaser.

Client 13 Donor Acquisition Mailing

A Teaser	17,532	375	170		11,215	66	6,575	4,641	27	0.970
B No teaser	17,527	375	143	27	35	60	6,573	1,948	14	0.816

Confidence level: 87%; unreliable, but probably worth retesting.

Tentative conclusion: teaser does not increase response.

Client 9 Donor Acquisition Mailing

A No teaser	30,000	346	182		5,173	28	10,370	-5,197	-29	0.606
B Teaser	30,000	346	176	6	38	26	10,370	-5,855	-33	0.586

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference; teaser does not increase response.

Client 6 Donor Acquisition Mailing

B Sticker, no teaser	18,313	415	117		3,467	30	7,600	-4,133	-35	0.639
C Sticker and teaser	18,308	415	130	13	31	30	7,598	-3,737	-29	0.710

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference.

Client 20 Donor Acquisition Mailing

A Teaser	30,004	368	264		3,686	14	11,045	-7,359	-28	0.880
B No teaser	29,996	368	304	40	48	14	11,042	-6,752	-22	1.013

Confidence level: 90%; probably significant, but you might want to retest.

Tentative conclusion: teaser does not increase response, but retest because teaser may lower response.

Client 21 Donor Acquisition Mailing

B Teaser	57,004	428	113		2,042	18	24,398	-22,356	-197	0.199
C No teaser	56,996	428	119	6	30	19	24,394	-22,150	-186	0.209

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference; teaser does not increase response.

Client 6 Member Renewal Series

A Teaser	12,234	960	1,250		51,732	41	11,745	39,987	32	10.220
B No teaser	12,233	960	1,104	146	97	42	11,744	34,728	31	9.025

Confidence level: 99%; it doesn't get much better than this!

Tentative conclusion: teaser increases response.

TABLE 23.1. Testing Outer Envelope Teasers, Continued

Sq. Root of									
Qty.	Cost/ 1,000	Total Gifts	Diff. Btwn. Gifts	Total Gifts × 2	Avg. Gift	Total Cost	Net Rev.	Acq. Cost/ Donor	% Resp.
Client 16 Donor Acquisition Mailing									
A Teaser	35,000	\$448	232		\$26	\$15,680	-\$9,651	-\$42	0.663%
B No teaser	35,000	448	246	14	25	15,680	-9,449	-38	0.703
Confidence level: less than 80%; you might as well flip a coin.									
Tentative conclusion: no statistically significant difference.									

BREs

For many years it was an article of faith among most direct mail fundraisers that the use of business reply envelopes (BREs), with the postage paid by the nonprofit organization, significantly lifted donor response. The additional expense—for the BRE permit, the half-dollar or so per-piece charge for postal processing, and the extra in-house processing work—was always assumed to be worth paying. It wasn't until recently that many fundraisers saw fit to question that premise.

Table 24.1 offers you the results of thirteen tests conducted in recent years for my clients. These tests examine the relative effectiveness of BREs and two kinds of reply envelopes with no prepaid postage: what my agency calls PSREs (place stamp reply envelopes) and NSREs (no stamp reply envelopes). A PSRE typically has a message in the upper-right-hand corner asking the donor to

supply the postage (for example, “Your First-Class Stamp Here, Please”); the NSRE has no such text. A fourth variation—the LSRE—is simply a self-addressed reply envelope to which a live first-class stamp has been applied.

As you’ll see in the following set of tests results, the BRE question is indeed an open one. Although many of these tests suggest that prospective donors respond equally well whether they must pay their own postage or not and that the postage-paid BRE probably represents money the nonprofit did not have to spend, there’s one notable exception, proving once again that you can’t take any outcome for granted.

TABLE 24.1.1. Testing BREs

		Sq. Root of Total				Acq. Cost/ Donor			
	Qty.	Cost/ 1,000	Total Gifts	Diff. Btwn. Gifts	Total Gifts × 2	Total Rev.	Avg. Gift	Total Cost	Net Rev.
Client 22 Donor Acquisition Test									
A BRE	40,000	\$419	416			\$9,205	\$22	\$16,760	-\$7,555
B PSRE	40,000	419	366	50	56	9,094	25	16,760	-7,666
Confidence level: 92%; probably significant, but you might want to retest.									
Tentative conclusion: BRE is an unnecessary additional expense.									
Client 16 Donor Acquisition Test									
A BRE	50,000	375	578			16,807	29	18,750	-1,943
C PSRE	50,000	375	576	2	68	17,293	30	18,750	-1,457
Confidence level: less than 80%; you might as well flip a coin.									
Tentative conclusion: no statistically significant difference; BRE does not increase response.									
Client 18 Donor Acquisition Test									
A BRE	18,000	398	123			5,097	41	7,164	-2,067
B PSRE	18,000	398	136	13	32	4,827	35	7,164	-2,337
Confidence level: less than 80%; you might as well flip a coin.									
Tentative conclusion: no statistically significant difference; BRE is an unnecessary additional expense and may lower response.									

[illegible]

Client 13 Donor Acquisition Test

A	NSRE	16,794	500	20		1,023	51	8,393	-7,370	-369	0.119	
B	BRE	16,794	500	17	3	12	881	52	8,393	-7,512	-443	0.101

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference; BRE is an unnecessary additional expense.

Client 6 Donor Acquisition Test

A	BRE	30,003	375	165		4,714	29	11,263	-6,549	-40	0.550	
B	PSRE	29,997	375	163	2	36	3,685	23	11,261	-7,576	-47	0.543

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference; BRE is an unnecessary additional expense.

Client 11 Donor Acquisition Test

A	BRE	30,002	322	427		6,310	15	9,674	-3,364	-8	1.423	
B	PSRE	29,998	322	398	29	57	6,071	15	9,673	-3,602	-9	1.326

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference; BRE is an unnecessary additional expense.

Client 11 Donor Acquisition Test

A	Rush BRE	30,003	327	438		6,198	14	9,799	-3,601	-8	1.460	
B	PSRE	29,997	327	429	9	59	6,100	14	9,797	-3,697	-9	1.430

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference.

Sq. Root of Total									
Qty.	Cost/ 1,000	Total Gifts	Diff. Btwn. Gifts	Total Gifts × 2	Avg. Gift	Total Cost	Net Rev.	Acq. Cost/ Donor	% Resp.
Client 9 Member Renewal Series									
A BRE, 1st notice	19,224	\$590	1,878		\$50	\$11,342	\$83,434	\$44	9.769%
B NSRE, 1st notice	19,217	590	1,842	122	51	11,338	83,465	45	9.585
Confidence level: less than 80%; you might as well flip a coin.									
Tentative conclusion: no statistically significant difference; BRE is an unnecessary additional expense.									
Client 7 Member Renewal Series									
A NSRE	14,142	590	384		56	8,344	13,156	34	2.715
B BRE	14,159	590	406	56	50	8,354	12,101	30	2.867
Confidence level: less than 80%; you might as well flip a coin.									
Tentative conclusion: no statistically significant difference; BRE is an unnecessary additional expense.									
Client 6 Member Renewal Series									
A PSRE, 1st notice	14,264	930	2,603		44	13,266	101,019	39	18.249
B LSRE, 1st notice	14,271	1,270	2,836	147	44	18,124	106,518	38	19.872
Confidence level: 99%; it doesn't get much better than this!									
Tentative conclusion: live stamp reply envelope lifts response and pays for itself.									

Letter Length

One of the oldest clichés in direct marketing is “long copy sells.” Ask any old-timer in the business to list the basic rules of direct response, and you’re likely to find these hallowed words on the list.

Cliché or not, this notion rests on solid evidence accumulated over several decades, and it’s risky to reject it out of hand. My own rule of thumb is, “Write as long a letter as you need to make the case for giving”—and lo and behold, for most organizations the case for giving a first gift usually fills several pages. (In contrast, the case for giving a renewal gift may often be made forcefully in only a few paragraphs.) But despite the weight of experience, it makes good sense to keep an open mind.

That’s exactly what my colleagues did in conducting the thirteen tests detailed in Table 25.1. The surprise was that often a

longer letter did not mean a greater response. But neither did it mean a lesser response. In fact letter length often did not produce enough variation to rise to the level of statistical significance. (However, whenever it did appear to make a difference, that difference, at least in these tests, favored the longer letter.) But review their results for yourself. Does this look like something your organization should be testing?

TABLE 25.1. Testing Letter Length

		Sq. Root of Total			Total			Avg.			Acq.		
	Qty.	Cost/ 1,000	Total Gifts	Diff. Btwn. Gifts	Total Gifts × 2	Total Rev.	Avg. Gift	Total Cost	Net Rev.	Donor	% Resp.		
Client 7 Donor Acquisition Mailing													
A 1-page letter	30,008	\$403	226			\$7,110	\$31	\$12,093	-\$4,983	-\$22	0.753%		
B 4-page letter	29,992	438	228	2	43	7,604	33	13,136	-5,532	-24	0.760		
Confidence level: less than 80%; you might as well flip a coin.													
Tentative conclusion: no statistically significant difference; four-page letter is an unnecessary additional expense.													
Client 6 Donor Acquisition Mailing													
A 4-page letter	60,000	383	684			17,391	25	22,980	-5,589	-8	1.140		
B 2-page letter	60,000	353	573	111	71	16,787	29	21,180	-4,393	-8	0.955		
Confidence level: 99%; it doesn't get much better than this!													
Tentative conclusion: four-page letter increases response.													
Client 8 House-File Mailing													
A 4-page letter	10,812	408	49			1,335	27	4,411	-3,076	-63	0.453		
C 2-page letter	10,809	392	51	2	20	2,000	39	4,237	-2,237	-44	0.472		
Confidence level: less than 80%; you might as well flip a coin.													
Tentative conclusion: no statistically significant difference; four-page letter is an unnecessary additional expense.													

	Sq. Root of	Total Gifts × 2	Avg. Gift	Total Cost	Net Rev.	Acq- Cost/ Donor	% Resp.
	Qty.	Cost/ 1,000	Total Gifts	Diff. Btwn. Gifts	Total Rev.		
Client 8 Donor Acquisition Mailing							
B 4-page letter	19,662	\$408	119		\$3,891	\$33	\$8,022
D 2-page letter	19,656	392	142	23	4,089	29	7,705
Confidence level: 84%; unreliable, but may be worth retesting.							
<i>Tentative conclusion: less expensive, two-page letter may increase response.</i>							
Client 6 Donor Acquisition Mailing							
A 1-page letter	40,000	418	203		5,526	27	16,720
C 4-page letter	40,000	440	244	41	6,800	28	17,600
Confidence level: 94%; probably significant, but you might want to retest.							
<i>Tentative conclusion: four-page letter may increase response, but retest.</i>							
Client 18 Donor Acquisition Mailing							
A 4-page letter	25,349	418	161		6,177	38	10,596
B 2-page letter	25,339	398	171	10	6,159	36	10,085
Confidence level: less than 80%; you might as well flip a coin.							
<i>Tentative conclusion: no statistically significant difference; two-page letter may be more cost effective.</i>							

Client 6 Donor Acquisition Mailing

A	1-page letter	40,000	362	231	6,150	27	14,480	-8,330	-36	0.578
B	4-page letter	40,000	374	313	82	47	14,960	-5,971	-19	0.783

Confidence level: 99%; it doesn't get much better than this!

Tentative conclusion: four-page letter increases response.

Client 6 Donor Acquisition Mailing

A 1-page letter	27,328	347	165		3,419	21	9,483	-6,064	-37	0.604	
B 4-page letter	27,325	372	147	18	35	3,967	27	10,165	-6,198	-42	0.538

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference; four-page letter is an unnecessary additional expense.

Client 24 Donor Acquisition Mailing

A 4-page letter	14,001	430	46		3,235	70	6,020	-2,785	-61	0.329	
B 2-page letter	13,999	420	42	4	19	1,186	28	5,880	-4,694	-112	0.300

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference; four-page letter is an unnecessary additional expense, but retest because of dramatic difference in average gift.

Client 24 Donor Acquisition Mailing

A 4-page letter	12,251	599	85		4,412	52	7,338	-2,926	-34	0.694	
B 2-page letter	12,248	589	81	4	26	2,931	36	7,214	-4,283	-53	0.661

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference; four-page letter is an unnecessary additional expense, but retest because of dramatic difference in average gift.

26

Gift Amounts

When I speak about the primary factors in the success or failure of a direct mail fundraising appeal, I note that the list is overwhelmingly the most important. Clearly, if you send a brilliant appeal to the wrong people, it should be no surprise that it bombs. But as I stressed in Part One, next in significance to the list is what marketers call the *offer*. In fundraising terms the key to the offer is usually the *ask*, the suggested minimum gift amount. And that amount, the topic of this chapter, can make all the difference in the world to the success of a mailing.

Sometimes.

Table 26.1 presents results from some sixty-eight tests conducted by my agency in recent years—every one of them intended to gauge the impact of variations in the suggested gift amount

or amounts, mostly in mailings for donor acquisition. The results of these tests were often intriguing enough to warrant at least a retest, and they should help you decide whether the amount of the suggested minimum ask is a question that warrants attention from your organization.

TABLE 26.1. Testing Gift Amounts

		Sq. Root of Total			Total			Avg.			Acq.		
	Qty.	Cost/ 1,000	Total Gifts	Diff. Btwn. Gifts	Total Gifts × 2	Total Rev.	Avg. Gift	Total Cost	Net Rev.	Donor	% Resp.		
Client 25 Donor Acquisition Mailing													
B \$30 ask	15,000	\$517	67			\$1,724	\$26	\$7,755	-\$6,031	-\$90	0.447%		
C \$35 ask	15,000	517	72	5	24	1,844	26	7,755	-5,911	-82	0.480		
Confidence level: less than 80%; you might as well flip a coin. Tentative conclusion: no statistically significant difference.													
Client 26 Donor Acquisition Mailing													
A \$30 ask	22,507	463	135			6,745	50	10,421	-3,676	-27	0.600		
B \$25 ask	22,508	463	166	31	35	7,966	48	10,421	-2,455	-15	0.738		
Confidence level: 92%; probably significant, but you might want to retest. Tentative conclusion: \$25 ask increases response; test with wider spread between ask amounts.													
Client 9 Donor Acquisition Mailing													
A \$20 ask (control)	50,008	394	576			12,339	21	19,703	-7,364	-13	1.152		
C \$100 ask	49,992	394	522	54	66	10,546	20	19,697	-9,151	-18	1.044		
Confidence level: 89%; unreliable, but probably worth retesting. Tentative conclusion: \$100 ask decreases response.													

Tentative conclusion: no statistically significant difference.

A	\$25 ask	40,703	390	211	7,304	35	15,874	-8,570	-41	0.518
B	\$30 ask	40,691	390	172	6,034	35	15,869	-9,835	-57	0.423

Tentative conclusion: \$25 ask increases response.

B	\$35 ask with blank option	26,678	414	232	5,559	24	11,045	-5,486	-24	0.870
C	\$35 ask with no blank option	26,653	414	157	75	39	11,034	-5,793	-37	0.589

Tentative conclusion: blank option is essential.

A	\$35, \$50, \$100 options	20,170	467	147	3,647	25	9,419	-5,772	-39	0.729
B	\$20, \$50, \$100 options	20,165	467	219	38	72	5,643	-3,774	-17	1.086

Tentative conclusion: starting with \$20 option increases response.

Sq. Root of Total												
Qty.	Cost/ 1,000	Total Gifts	Diff. Btwn. Gifts	Total Gifts × 2	Total Rev.	Avg. Gift	Total Cost	Net Rev.	Acq. Cost/ Donor	% Resp.		
Client 1 Donor Acquisition Mailing												
A	\$50, will be matched	24,434	\$460	120			\$4,644	\$39	\$11,240	−\$6,596	−\$55	0.491%
B	\$20, will be matched	24,425	460	142	22	32	5,659	40	11,236	−5,577	−39	0.581
Confidence level: 86%; unreliable, but probably worth retesting.												
Tentative conclusion: \$20 ask may generate higher response; retest.												
Client 10 Donor Acquisition Mailing												
A	\$15, \$25, \$50 options	40,000	376	367			7,275	20	15,040	−7,765	−21.16	0.918
B	\$15, \$24, \$48, \$72 options	40,000	376	337	30	53	7,234	21	15,040	−7,806	−23.16	0.843
Confidence level: less than 80%; you might as well flip a coin.												
Tentative conclusion: no statistically significant difference; can stick with “traditional” gift amounts.												
Client 9 Donor Acquisition Mailing												
A	Blue RD, no ask circled (control)	30,000	448	223			4,868	22	13,440	−8,572	−38	0.743
C	Blue RD, \$30 ask circled	30,000	448	173	50	40	4,850	28	13,440	−8,590	−50	0.577
Confidence level: 99%; it doesn’t get much better than this!												
Tentative conclusion: circled \$30 on RD decreases response.												

Client 7 Donor Acquisition Mailing

A \$26.75 ask	23,598	430	170		5,112	30	10,147	-5,035	-30	0.720
B \$35 ask	23,594	430	171	1	37	31	10,145	-4,769	-28	0.725

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference.

Client 7 Donor Acquisition Mailing

A \$26.75 ask with blank option	23,598	430	170		5,112	30	10,147	-5,035	-30	0.720
C \$26.75 ask with no blank option	23,586	430	164	6	37	32	10,142	-4,852	-30	0.695

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference.

Client 15 Donor Acquisition Mailing

A 8.5" x 11" RD, no ask circled	30,000	373	235		6,354	27	11,190	-4,836	-21	0.783
C 8.5" x 11" RD, \$50 ask circled	30,000	373	169	66	40	36	11,190	-5,181	-31	0.563

Confidence level: 99%; it doesn't get much better than this!

Tentative conclusion: response device without circled ask increases response.

Sq. Root of											
Total											
Qty.	Cost/ 1,000	Total Gifts	Diff. Btwn. Gifts	Total Gifts × 2	Avg. Gift	Total Cost	Net Rev.	Acq. Cost/ Donor	% Resp.		
Client 9 Donor Acquisition Mailing											
A \$20, \$30, and blank options	30,000	\$410	252		\$6,004	\$24	\$12,300	−\$6,296	−\$25	0.840%	
B \$20, \$30 options, no blank option	30,000	410	242	10	44	6,427	27	12,300	−5,873	−24	0.807
Confidence level: less than 80%; you might as well flip a coin.											
Tentative conclusion: no statistically significant difference.											
Next entry continues this test.											
Client 9 Donor Acquisition Mailing											
A \$20, \$30, and blank options	30,000	410	252		6,004	24	12,300	−6,296	−25	0.840	
C \$15, \$25, and blank options	30,000	410	263	11	45	6,394	24	12,300	−5,906	−22	0.877
Confidence level: less than 80%; you might as well flip a coin.											
Tentative conclusion: no statistically significant difference.											

Client 7 Donor Acquisition Mailing

A	\$35 ask (control)	30,000	390	157		5,012	32	11,700	-6,688	-43	0.523
B	\$26.75 ask circled	30,000	390	180	23	37	31	11,700	-6,048	-34	0.600

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference.

Next entry continues this test.**Client 7 Donor Acquisition Mailing**

A	\$35 ask (control)	30,000	390	157		5,012	32	11,700	-6,688	-43	0.523
C	\$26.75 ask not circled	30,000	390	215	58	39	36	11,700	-4,026	-19	0.717

Confidence level: 99%; it doesn't get much better than this!

Tentative conclusion: \$26.75 ask increases response.

Next entry continues this test.**Client 7 Donor Acquisition Mailing**

B	\$26.75 ask circled	30,000	390	180		5,652	31	11,700	-6,048	-34	0.600
C	\$26.75 ask not circled	30,000	390	215	35	40	36	11,700	-4,026	-19	0.717

Confidence level: 92%; probably significant, but you might want to retest.

Tentative conclusion: circled amount may lower response; retest.

164

Client 7 Donor Acquisition Mailing

A \$35 ask (control)	30,000	389	275		7,918	29	11,670	-3,752	-14	0.917	
B \$20 ask	30,000	389	281	6	47	6,636	24	11,670	-5,034	-18	0.937

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference.

Client 15 Donor Acquisition Mailing

A \$10 ask	30,000	333	246		5,911	24	9,990	-4,079	-17	0.820	
B \$20 ask	30,000	333	184	62	41	5,298	29	9,990	-4,692	-26	0.613

Confidence level: 99%; it doesn't get much better than this!

Tentative conclusion: \$10 ask increases response.

Next entry continues this test.**Client 15 Donor Acquisition Mailing**

A \$10 ask	30,000	333	246		5,911	24	9,990	-4,079	-17	0.820	
C \$20-\$100 options, \$100 circled	30,000	333	186	60	42	5,974	32	9,990	-4,016	-22	0.620

Confidence level: 99%; it doesn't get much better than this!

Tentative conclusion: \$10 ask increases response.

Client 8 Donor Acquisition Mailing

Client 9 Donor Acquisition Mailing

A	\$20 ask (control)	30,000	417	255		6,153	24	12,510	-6,357	-25	0.850
B	\$21.50 ask	30,000	417	229	26	44	22	12,510	-7,532	-33	0.763

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference.

Next entry continues this test.

Client 9 Donor Acquisition Mailing

A	\$20 ask (control)	30,000	417	255		6,153	24	12,510	-6,357	-25	0.850
C	\$35 ask	30,000	417	234	21	44	23	12,510	-7,030	-30	0.780

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference.

Next entry continues this test.

Client 9 Donor Acquisition Mailing

B	\$21.50 ask	30,000	417	229		4,978	22	12,510	-7,532	-33	0.763
C	\$35 ask	30,000	417	234	5	43	23	12,510	-7,030	-30	0.780

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference.

Next entry continues this test.

Next entry continues this test.

Client 7 Donor Acquisition Mailing

B \$25 ask	30,000	357	262			6,016	23	10,710	-4,694	-18	0.873
C \$20 ask	30,000	357	310	48	48	7,341	24	10,710	-3,369	-11	1.033

Confidence level: 95%; statistically significant.

Tentative conclusion: \$20 ask increases response.

Client 9 Donor Acquisition Mailing

A \$26 ask (control)	30,000	376	192			4,713	25	11,280	-6,567	-34	0.640
B \$20 ask (control)	30,000	376	239	47	42	5,198	22	11,280	-6,082	-25	0.797

Confidence level: 97%; statistically significant.

Tentative conclusion: \$20 ask increases response.

Next entry continues this test.

Client 9 Donor Acquisition Mailing

A \$26 ask (control)	30,000	376	192			4,713	25	11,280	-6,567	-34	0.640
C \$26 ask (new)	30,000	376	174	18	38	4,683	27	11,280	-6,597	-38	0.580

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference.

TABLE 26.1. Testing Gift Amounts, Continued

Sq. Root of Total												
Qty.	Cost/ 1,000	Total Gifts	Diff. Btwn. Gifts	Total Gifts × 2	Total Rev.	Avg. Gift	Total Cost	Net Rev.	Acq. Cost/ Donor	% Resp.		
Next entry continues this test.												
Client 9 Donor Acquisition Mailing												
B	\$20 ask (control)	30,000	\$376	239			\$5,198	\$22	\$11,280	-\$6,082	-\$25	0.797%
D	\$20 ask (new)	30,000	376	211	28	42	4,485	21	11,280	-6,795	-32	0.703
Confidence level: 81%; unreliable, but may be worth retesting.												
Tentative conclusion: no statistically significant difference.												
Next entry continues this test.												
Client 9 Donor Acquisition Mailing												
C	\$26 ask (new)	30,000	376	174			4,683	27	11,280	-6,597	-38	0.580
D	\$20 ask (new)	30,000	376	211	37	39	4,485	21	11,280	-6,795	-32	0.703
Confidence level: 94%; probably significant, but you might want to retest.												
Tentative conclusion: \$20 ask may increase total gifts but may lower average gift, retest.												

Client 10 Donor Acquisition Mailing

A \$25 ask	70,000	364	433			11,205	26	25,480	-14,275	-33	0.619
B \$15 ask	70,000	364	490	57	61	10,575	22	25,480	-14,905	-30	0.700

Confidence level: 94%; probably significant, but you may want to retest

Tentative conclusion: \$15 ask may increase total gifts but may lower average gift; retest.

Client 5 Donor Acquisition Mailing

A \$25 ask	30,004	410	145			4,547	31	12,302	-7,755	-53	0.483
B \$35 ask	29,996	410	159	14	35	5,408	34	12,298	-6,890	-43	0.530

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference.

Client 8 Donor Acquisition Mailing

A \$20 ask	25,000	349	215			5,626	26	8,725	-3,099	-14	0.860
B \$35 ask	25,000	349	193	22	40	5,296	27	8,725	-3,429	-18	0.772

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference.

Client 23 Donor Acquisition Mailing

A \$20 ask	25,358	560	511			13,049	26	14,201	-1,152	-2	2.015
B \$25 ask	25,357	560	505	6	64	11,907	24	14,201	-2,294	-5	1.992

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference.

Next entry continues this test.

Next entry continues this test.

Client 19 Donor Acquisition Mailing

B	\$35 ask	30,000	371	273		8,468	31	11,130	-2,662	-10	0.910	
C	\$25 ask	30,000	371	242	31	45	7,514	31	11,130	-3,616	-15	0.807

Confidence level: 83%; unreliable, but may be worth retesting.

Tentative conclusion: no statistically significant difference, but \$35 ask may increase income.

Client 9 Donor Acquisition Mailing

A	\$26.75 ask with labels	30,000	394	225		6,154	27	11,820	-5,666	-25	0.750	
B	\$25 ask with labels	30,000	394	236	11	43	4,754	20	11,820	-7,066	-30	0.787

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference.

Client 23 Donor Acquisition Mailing

A	\$20 ask with \$100 max.	21,985	488	424	9,869	23	10,732	-863	-2	1.929	
B	\$19.98 ask with \$100 max.	21,976	488	463	60	10,764	23	10,727	37	0	2.107

Confidence level: 81%; unreliable, but may be worth retesting.

Tentative conclusion: no statistically significant difference.

Client 6 Donor Acquisition Mailing

A \$25 ask with short letter	27,328	347	165	3,419	21	9,483	-6,064	-37	0.604
D \$15 ask with short letter	27,321	346	146	2,907	20	9,453	-6,546	-45	0.534

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference.

Client 5 Donor Acquisition Mailing

A \$25 ask	20,239	428	203	5,842	29	8,662	-2,820	-14	1.003
B \$35 ask	20,232	428	194	6,606	34	8,659	-2,053	-11	0.959

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference.

Next entry continues this test.

Client 5 Donor Acquisition Mailing

A \$25 ask	20,239	428	203	5,842	29	8,662	-2,820	-14	1.003
C \$50 ask	20,229	428	160	5,776	36	8,658	-2,882	-18	0.791

Confidence level: 97%; statistically significant.

Tentative conclusion: \$25 ask increases response.

Sq. Root of Total											
		Cost/ 1,000	Total Gifts	Diff. Btwn. Gifts	Total Gifts × 2	Total Rev.	Avg. Gift	Total Cost	Net Rev.	Acq. Cost/ Donor	% Resp.
Next entry continues this test.											
Client 5 Donor Acquisition Mailing											
B	\$35 ask	20,232	\$428	194		\$6,606	\$34	\$8,659	−\$2,053	−\$11	0.959%
C	\$50 ask	20,229	428	160	34	5,776	36	8,658	−2,882	−18	0.791
Confidence level: 93%; probably significant, but you might want to retest.											
Tentative conclusion: \$35 ask may increase response.											
Client 28 Donor Acquisition Mailing											
A	\$30 ask	25,000	422	135		4,024	30	10,557	−6,533	−48	0.540
B	\$50 ask	25,000	422	110	25	2,897	26	10,557	−7,660	−70	0.440
Confidence level: 89%; unreliable, but probably worth retesting.											
Tentative conclusion: no statistically significant difference, but retest because of difference in revenue.											
Client 9 Donor Acquisition Mailing											
A	\$27.75 ask (control)	30,000	357	191		5,202	27	10,710	−5,508	−29	0.637
B	\$50 ask	30,000	357	178	13	4,524	25	10,710	−6,186	−35	0.593
Confidence level: less than 80%; you might as well flip a coin.											
Tentative conclusion: no statistically significant difference.											

Next entry continues this test.

Client 9 Donor Acquisition Mailing

A	\$27.75 ask (control)	30,000	357	191		5,202	27	10,710	-5,508	-29	0.637	
C	\$27.75 ask (null test)	30,000	357	218	27	40	6,169	28	10,710	-4,541	-21	0.727

Confidence level: 81%; unreliable, but may be worth retesting.

Tentative conclusion: no statistically significant difference.

Next entry continues this test.

Client 9 Donor Acquisition Mailing

B	\$50 ask	30,000	357	178		4,524	25	10,710	-6,186	-35	0.593	
C	\$27.75 ask (control)	30,000	357	218	40	40	6,169	28	10,710	-4,541	-21	0.727

Confidence level: 95%; statistically significant.

Tentative conclusion: \$27.75 ask increases response.

Client 6 Donor Acquisition Mailing

A \$25 ask	30,000	395	192		10,173	53	11,850	-1,677	-9	0.640	
B \$15 ask	30,000	395	219	27	41	5,148	24	11,850	-6,702	-31	0.730

Confidence level: 81%; unreliable, but may be worth retesting.

Tentative conclusion: no statistically significant difference, but retest because of dramatic difference in average gift.

Sq. Root of Total											
	Qty.	Cost/ 1,000	Total Gifts	Diff. Btwn. Gifts	Total Gifts × 2	Total Rev.	Avg. Gift	Total Cost	Net Rev.	Acq. Cost/ Donor	% Resp.
Client 8 Donor Acquisition Mailing											
A	\$25 ask	26,874	\$371	150		\$4,513	\$30	\$9,970	−\$5,457	−\$36	0.558%
B	\$30 ask	26,869	371	118	32	3,657	31	9,968	−6,311	−53	0.439
Confidence level: 94%; probably significant, but you might want to retest. Tentative conclusion: \$25 ask may increase response.											
Next entry continues this test.											
Client 8 Donor Acquisition Mailing											
A	\$25 ask	26,874	371	150		4,513	30	9,970	−5,457	−36	0.558
C	\$20 ask	26,859	371	187	37	5,280	28	9,965	−4,685	−25	0.696
Confidence level: 95%; statistically significant. Tentative conclusion: \$20 ask increases response.											
Next entry continues this test.											
Client 8 Donor Acquisition Mailing											
B	\$30 ask	26,869	371	118		3,657	31	9,968	−6,311	−53	0.439
C	\$20 ask	26,859	371	187	69	5,280	28	9,965	−4,685	−25	0.696
Confidence level: 99%; it doesn't get much better than this! Tentative conclusion: \$20 ask increases response.											

Client 9 Donor Acquisition Mailing

A \$20 ask (control)	50,005	397	309		8,590	28	19,875	-11,285	-37	0.618
B \$26.75 ask circled	50,002	397	425	116	54	27	19,874	-8,434	-20	0.850

Confidence level: 99%; it doesn't get much better than this!

Tentative conclusion: \$26.75 ask with circle increases response.

Client 17 Donor Acquisition Mailing

A \$25 ask	14,079	630	71		1,878	26	8,870	-6,992	-98	0.504
B \$15 ask	14,077	630	96	25	26	28	8,869	-6,162	-64	0.682

Confidence level: 94%; probably significant, but you might want to retest.

Tentative conclusion: \$15 ask increases response.

Client 5 Donor Acquisition Mailing

A \$25 ask	23,004	450	253		8,971	35	10,352	-1,381	-5	1.100
B \$20 ask	23,001	450	247	6	45	25	10,350	-4,141	-17	1.074

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference.

Next entry continues this test.**Client 5 Donor Acquisition Mailing**

A \$25 ask	23,004	450	253		8,971	35	10,352	-1,381	-5	1.100
C \$30 ask	22,995	450	263	10	45	30	10,348	-2,340	-9	1.144

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference.

180

Gift Strings

Observant readers of direct mail fundraising appeals will have noticed that on most appeals the minimum suggested gift amount is placed first in a series (for example, “I am enclosing my tax-deductible gift in the amount of: ☐ \$25 ☐ \$35 ☐ \$50 ☐ \$100 ☐ \$___”). From time to time, however, the order is reversed (for example, “I am enclosing my tax-deductible gift in the amount of: ☐ \$100 ☐ \$50 ☐ \$35 ☐ \$25 ☐ \$___”).

Is there any logic to this reversal? Does it affect the rate of response or the average gift? Presumably, the answer to at least one of these questions is yes. Otherwise, assuming that direct mail fundraisers are rational and cautious people who test such things, we would never see the order reversed, would we?

Table 27.1 offers the results of four tests of this proposition—clearly not enough to establish a pattern but thought provoking nonetheless, particularly in the average gift column.

TABLE 27.1. Testing Gift Strings

Sq. Root of Total											
	Qty.	Cost/ 1,000	Total Gifts	Diff. Btwn. Gifts	Total Gifts × 2	Avg. Gift	Total Cost	Net Rev.	Acq. Cost/ Donor	% Resp.	
Client 8 Donor Acquisition Mailing											
A	Ascending gift string	33,998	\$425	163		\$35	\$14,449	−\$8,672	−\$53	0.479%	
B	Descending gift string	33,987	425	162	1	36	6,819	42	14,444	−7,625	−47
Confidence level: less than 80%; you might as well flip a coin.											
Tentative conclusion: no statistically significant difference; almost identical response; descending order has higher average gift.											
Client 3 Donor Acquisition Mailing											
A	Ascending gift string	40,008	400	316		38	16,003	−4,142	−13	0.790	
B	Descending gift string	39,992	400	346	30	51	13,558	39	15,997	−2,439	−7
Confidence level: less than 80%; you might as well flip a coin.											
Tentative conclusion: no statistically significant difference.											

Client 13 Donor Acquisition Mailing

A	Descending gift string	14,778	542	74		8,729	118	8,010	719	10	0.501
B	Ascending gift string	14,770	542	67	7	24	5,875	88	-2,130	-32	0.454

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference, but retest because of difference in average gift.

Client 13 Donor Acquisition Mailing

A	Ascending gift string	15,258	510	90		11,905	132	7,782	4,123	46	0.590
B	Descending gift string	15,263	509	99	9	27	7,322	74	-447	-5	0.649

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference, but retest because of big difference in average gift.

Credit Card Options

Traditionally, direct mail fundraising in the United States has made use of the pervasiveness of checking accounts as a method of paying bills and making cash gifts. Before the last two decades, it was unusual to encounter a fundraising appeal that offered to accept gifts made by credit card. Even today, though the practice has become more common, the credit card option appears on only a minority of the reply devices for appeals.

Think about whether it makes sense for you to offer your donors a credit card option. Review the results of the eight tests in this chapter. Then test for yourself.

Sq. Root of Total										
Qty.	Cost/ 1,000	Total Gifts	Diff. Btwn. Gifts	Total Gifts × 2	Total Rev.	Avg. Gift	Total Cost	Net Rev.	Acq. Cost/ Donor	% Resp.
Client 15 Donor Acquisition Mailing										
A&B	Credit card option	124,473	\$391	937	\$31,645	\$34	\$48,669	−\$17,024	−\$18	0.753%
C&D	No credit card option	124,446	391	911	26	30,741	34	48,658	−20	0.732
Confidence level: less than 80%; you might as well flip a coin. Tentative conclusion: no statistically significant difference; credit card option does not increase response.										
Client 22 Donor Acquisition Mailing										
A	Credit card option	50,000	397	348	9,069	26	19,850	−10,781	−31	0.696
B	No credit card option	50,000	394	369	21	9,952	27	19,700	−26	0.738
Confidence level: less than 80%; you might as well flip a coin. Tentative conclusion: no statistically significant difference; credit card option does not increase response.										
Client 6 Donor Acquisition Mailing										
A	No credit card option	60,000	408	560	11,762	21	24,480	−12,718	−23	0.933
B	Credit card option	60,000	408	551	9	11,865	22	24,480	−23	0.918
Confidence level: less than 80%; you might as well flip a coin. Tentative conclusion: no statistically significant difference; credit card option does not increase response.										

Client 7 Donor Acquisition Mailing

A	No credit card option	30,000	364	256		8,056	31	10,920	-2,864	-11.19	0.853
B	Credit card option	30,000	364	267	11	46	7,886	30	10,920	-3,034	-11.36

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference; credit card option does not increase response.

Client 27 Donor Acquisition Mailing

A	No credit card option	25,000	411	316		8,855	28	10,275	-1,420	-4	1.264
B	Credit card option	25,000	411	283	33	49	8,546	30	10,275	-1,729	-6

Confidence level: 82%; unreliable, but may be worth retesting.

Tentative conclusion: credit card option does not increase response.

Client 16 Donor Acquisition Mailing

A	No credit card option (control)	30,000	388	335		10,762	32	11,640	-878	-3	1.117
B	Credit card option	30,000	388	325	10	51	9,244	28	11,640	-2,396	-7

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference; credit offer does not increase response.

Client 23 Donor Acquisition Mailing

C	Credit card option	32,500	472	754		21,857	29	15,349	6,508	9	2.320
A	No credit card option	32,500	472	848	94	80	23,492	28	15,349	8,143	10

Confidence level: 98%; we have a winner!

Tentative conclusion: credit card option decreases response.

TABLE 28.1. Testing Credit Card Options, Continued

Sq. Root of Total												
Qty.	Cost/ 1,000	Total Gifts	Diff. Btwn. Gifts	Total Gifts × 2	Avg. Gift	Total Cost	Net Rev.	Acq. Cost/ Donor	% Resp.			
Client 7 Member Renewal Series												
A	Credit card option	14,702	\$590	479		\$27,485	\$57	\$8,674	\$18,811	\$39	3,258%	
B	No credit card option	14,507	590	505	26	63	35,551	70	8,559	26,992	53	3,481
Confidence level: less than 80%; you might as well flip a coin.												
Tentative conclusion: no statistically significant difference, but retest because of dramatic difference in average gift.												

29

Miscellaneous Tests

If you have been wondering about a particular type of test, chances are you'll find an example of it in Table 29.1. Featured here are ninety-six separate tests my colleagues and I have conducted in recent years in addition to the tests shown in the previous nine tables. Here you'll see package tests galore—pitting one package against another—as well as a plethora of tests probing such questions as the value of petitions, pledges, surveys, and premiums; the best format for reply devices; and the reaction when a nonprofit changes its logo—questions likely to be of considerable importance to your organization at some time in its quest for new donors. As you devise your own test, you will have the guidance of at least a few of the results experienced by others.

Many of the tests involve head-to-head comparisons between two packages that are not described in any specific way. Don't fret:

the particular differences between any two of these packages are far less important than the fact that they are *different* packages. Reviewing the results of these tests will give you a sense of the relative value of testing an entirely new package against a control as compared to testing lesser variations such as inserts, suggested minimum gift amounts, or postage rates.

TABLE 29.1. Testing Other Package Elements: Logos, Surveys, Reply Devices, Typefaces, Booklets, and More

Sq. Root of Total													
		Cost/ 1,000		Total Gifts	Diff. Btwn. Gifts	Total Gifts × 2	Total Rev.	Avg. Gift	Total Cost	Net Rev.	Acq. Cost/ Donor	%	
Qty.													
Client 6 Donor Acquisition Mailing													
A		Citizen petition	50,006	\$397	497		\$13,254	\$27	\$19,852	-\$6,598	-\$13	0.994%	
B		Citizen pledge	49,994	397	396	101	60	13,779	35	19,848	-6,069	-15	0.792
Confidence level: 99%; it doesn't get much better than this!													
Tentative conclusion: citizen petition increases response.													
Client 2 Donor Acquisition Mailing													
A		Control pkg. with no petition	25,000	346	286				8,212	29	8,650	-2	1.144
C		Pkg. 1 with petition	25,000	386	315	29	49	8,713	28	9,650	-937	-3	1.260
Confidence level: less than 80%; you might as well flip a coin.													
Tentative conclusion: no statistically significant difference.													
Next entry continues this test.													
Client 2 Donor Acquisition Mailing													
B		Pkg. 1 with no petition	25,000	346	282			7,808	28	8,650	-842	-3	1.128
C		Pkg. 1 with petition	25,000	386	315	33	49	8,713	28	9,650	-937	-3	1.260
Confidence level: 82%; unreliable, but may be worth retesting.													
Tentative conclusion: petition does not increase response.													

TABLE 29.1. Testing Other Package Elements, Continued

Sq. Root of Total										
Qty.	Cost/ 1,000	Total Gifts	Diff. Btwn. Gifts	Total Gifts × 2	Total Rev.	Avg. Gift	Total Cost	Net Rev.	Acq. Cost/ Donor	% Resp.
Client 6 Donor Acquisition Mailing										
A Petition 1	20,000	\$364	154		\$5,948	\$39	\$7,280	−\$1,332	−\$9	0.770%
B Survey 1	20,000	364	109	45	3,102	28	7,280	−4,178	−38	0.545
Confidence level: 99%; it doesn't get much better than this!										
Tentative conclusion: petition increases response.										
Next entry continues this test.										
Client 6 Donor Acquisition Mailing										
C Petition 2	20,000	364	142		4,211	30	7,280	−3,069	−22	0.710
D Survey 2	20,000	364	114	28	3,066	27	7,280	−4,214	−37	0.570
Confidence level: 92%; probably significant, but you might want to retest.										
Tentative conclusion: petition increases response.										
Next entry continues this test.										
Client 6 Donor Acquisition Mailing										
A Petitions 1 and 2 (2 tests)	40,000	364	296		10,159	34	14,560	−4,401	−15	0.740
B Surveys 1 and 2 (2 tests)	40,000	364	223	73	6,168	28	14,560	−8,392	−38	0.558
Confidence level: 99%; it doesn't get much better than this!										
Tentative conclusion: petitions generate higher response than surveys.										

Client 16 Donor Acquisition Mailing

A Survey (control)	41,674	414	357	9,704	27	17,253	-7,549	-21	0.857		
B Survey with calendar offer	41,673	414	348	9	53	10,842	31	17,253	-6,411	-18	0.835

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference; calendar is an unnecessary additional expense.

Next entry continues this test.**Client 16 Donor Acquisition Mailing**

A Survey (control)	41,674	414	357	9,704	27	17,253	-7,549	-21	0.857
C Test pkg. 1 (no survey)	41,672	414	161	4,990	31	17,252	-12,262	-76	0.386

Confidence level: 99%; it doesn't get much better than this!

Tentative conclusion: survey package increases response.

Next entry continues this test.**Client 16 Donor Acquisition Mailing**

A Survey (control)	41,674	414	357		9,704	27	17,253	-7,549	-21	0.857	
D Calendar offer (no survey)	41,671	414	310	47	52	10,472	34	17,252	-6,780	-22	0.744

Confidence level: 92%; probably significant, but you might want to retest.

Tentative conclusion: survey package increases response.

Sq. Root of Total										
Qty.	Cost/ 1,000	Total Gifts	Diff. Btwn. Gifts	Total Gifts × 2	Total Rev.	Avg. Gift	Total Cost	Net Rev.	Acq. Cost/ Donor	% Resp.
Next entry continues this test.										
Client 16 Donor Acquisition Mailing										
B Survey with calendar offer	41,673	\$414	348		\$10,842	\$31	\$17,253	-\$6,411	-\$18	0.835%
C Test pkg. 1 (no survey)	41,672	414	161	45	4,990	31	17,252	-12,262	-76	0.386
Confidence level: 99%; it doesn't get much better than this!										
Tentative conclusion: survey with calendar offer increases response.										
Next entry continues this test.										
Client 16 Donor Acquisition Mailing										
B Survey with calendar offer	41,673	414		348	10,842	31	17,253	-6,411	-18	0.835
D Calendar offer (no survey)	41,671	414	310	51	10,472	34	17,252	-6,780	-22	0.744
Confidence level: 87%; unreliable, but probably worth retesting.										
Tentative conclusion: no real conclusion.										

Next entry continues this test.

Client 16 Donor Acquisition Mailing

C Test pkg. 1 (no survey)	41,672	414	161		4,990	31	17,252	-12,262	-76	0.386	
D Calendar offer (no survey)	41,671	414	310	149	43	10,472	34	17,252	-6,780	-22	0.744

Confidence level: 99%; it doesn't get much better than this!

Tentative conclusion: calendar package increases response.

Client 16 Donor Acquisition Mailing

A No survey	25,000	428	383		12,338	32	10,700	1,638	4	1.532	
B Survey (control)	25,000	428	535	152	61	15,832	30	10,700	5,132	10	2.140

Confidence level: 99%; it doesn't get much better than this!

Tentative conclusion: survey increases response.

Next entry continues this test.

Client 16 Donor Acquisition Mailing

B Survey (control)	25,000	428	383		12,338	32	10,700	1,638	4	1.532	
C Test pkg. 2 (no survey)	25,000	428	244	139	50	8,425	35	10,700	-2,275	-9	0.976

Confidence level: 99%; it doesn't get much better than this!

Tentative conclusion: test package decreases response.

Client 11 Donor Acquisition Mailing

A	Survey (control)	30,000	338	459		7,596	17	10,136	-2,540	-6	1,530	
B	No survey	30,000	338	440	19	60	7,328	17	10,136	-2,808	-6	1,467

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference.

Client 11 Donor Acquisition Mailing

A	Survey (control)	30,000	345	601	9,735	16	10,337	-603	-1	2,003
B	No survey	30,000	345	455	146	65	7,527	17	10,337	-2,810
									-6	1,517

Confidence level: 99%; it doesn't get much better than this!

Tentative conclusion: survey increases response.

Client 6 Donor Acquisition Mailing

A	Pledge (control)	40,007	380	241		7,083	29	15,203	-8,120	-34	0.602	
B	Note card	39,993	500	187	54	41	5,208	28	19,997	-14,789	-79	0.468

Confidence level: 99%; it doesn't get much better than this!

Tentative conclusion: pledge control package increases response.

Client 7 Donor Acquisition Mailing

A	RD with info on back (control)	30,001	387	265		6,482	24	11,610	-5,128	-19	0.883	
C	RD with blank back	30,001	387	269	4	46	6,497	24	11,610	-5,113	-19	0.897

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference.

TABLE 29.1. Testing Other Package Elements, Continued

Sq. Root of Total									
Qty.	Cost/ 1,000	Total Gifts	Diff. Btwn. Gifts	Total Gifts × 2	Avg. Gift	Total Cost	Net Rev.	Acq. Cost/ Donor	% Resp.
Client 14 Donor Acquisition Mailing									
A RD with info on back (control)	20,000	\$491	212		\$23	\$9,820	−\$4,903	−\$23	1.060%
C RD with blank back	20,000	491	224	42	25	9,820	−4,168	−19	1.120
Confidence level: less than 80%; you might as well flip a coin.									
Tentative conclusion: no statistically significant difference.									
Client 1 Donor Acquisition Mailing									
B RD with design 1	19,237	427	88		34	8,214	−5,219	−59	0.457
C RD with design 2	19,231	427	65	25	35	8,212	−5,922	−91	0.338
Confidence level: 93%; probably significant, but you might want to retest.									
Tentative conclusion: RD with design 1 may increase response, but retest.									
Client 9 Donor Acquisition Mailing									
A RD with copy on back	30,000	404	356		24	12,120	−3,672	−10	1.187
B RD with blank back	30,000	404	396	55	27	12,120	−1,504	−4	1.320
Confidence level: 85%; unreliable, but probably worth retesting.									
Tentative conclusion: RD with no copy on back may increase response, but retest.									

Client 9 Donor Acquisition Mailing

A Blue RD, ask not circled	30,000	448	223		4,868	22	13,440	-8,572	-38	0.743
B Red RD, ask not circled	30,000	448	229	6	43	25	13,440	-7,792	-34	0.763

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference; use RD of either color.

Client 15 Donor Acquisition Mailing

A RD with ask not circled	30,000	373	235		6,354	27	11,190	-4,836	-21	0.783
C RD with \$50 ask circled	30,000	373	169	66	40	36	11,190	-5,181	-31	0.563

Confidence level: 99%; it doesn't get much better than this!

Tentative conclusion: RD without circled amount increases response.

Client 7 Donor Acquisition Mailing

A RD with no note	30,000	453	172		4,624	27	13,590	-8,966	-52	0.573
B RD with note 1	30,000	453	164	8	37	25	13,590	-9,465	-58	0.547

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference.

Sq. Root of Total							
Qty.	Cost/ 1,000	Total Gifts	Diff. Btwn. Gifts	Total Gifts × 2	Total Rev.	Avg. Gift	Total Cost
					Net Rev.		Acq- Cost/ Donor
							% Resp.
Client 7 Donor Acquisition Mailing							
A RD with no note	30,000	\$453	172		\$4,624	\$27	\$13,590
C RD with note 2	30,000	453	159	36	4,110	26	13,590
Confidence level: less than 80%; you might as well flip a coin.							
Tentative conclusion: no statistically significant difference.							
Client 7 Donor Acquisition Mailing							
B RD with note 1	30,000	453	164		4,125	25	13,590
C RD with note 2	30,000	453	159	36	4,110	26	13,590
Confidence level: less than 80%; you might as well flip a coin.							
Tentative conclusion: no statistically significant difference.							
Client 9 Donor Acquisition Mailing							
C #10 RD, 1-side	30,000	487	193		5,889	31	14,610
D #10 RD, 2-side	30,000	487	193	39	6,155	32	14,610
Confidence level: less than 80%; you might as well flip a coin.							
Tentative conclusion: no statistically significant difference.							

Client 7 Donor Acquisition Mailing

C RD with portrait 1	30,000	348	108		3,350	31	10,440	-7,090	-66	0.360
D RD with portrait 2	30,000	348	89	19	28	32	10,440	-7,615	-86	0.297

Confidence level: 82%; unreliable, but may be worth retesting.

Tentative conclusion: no statistically significant difference.

Client 23 Donor Acquisition Mailing

A 8.5" x 14" RD (control)	14,065	504	316		9,170	29	7,093	2,077	7	2.247
B 8.5" x 11" RD	14,058	504	330	14	51	27	7,090	1,740	5	2.347

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference.

Client 6 Donor Acquisition Mailing

A RD with no ballot (control)	30,003	354	147		3,984	27	10,621	-6,637	-45	0.490
B RD with ballot	29,997	354	164	17	35	24	10,619	-6,646	-41	0.547

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference.

Client 9 Donor Acquisition Mailing

A RD with design 1 (control)	50,006	380	449		13,321	30	19,002	-5,681	-13	0.898
B RD with design 2	50,006	380	597	148	65	19	19,002	-7,775	-13	1.194

Confidence level: 99%; it doesn't get much better than this!

Tentative conclusion: RD with design 2 increases response.

TABLE 29.1. Testing Other Package Elements, Continued

Sq. Root of Total										
Qty.	Cost/ 1,000	Total Gifts	Diff. Btwn. Gifts	Total Gifts × 2	Total Rev.	Avg. Gift	Total Cost	Net Rev.	Acq. Cost/ Donor	% Resp.
Client 6 Donor Acquisition Mailing										
A RD with design 1 (control)	30,003	\$419	100		\$2,058	\$21	\$12,571	−\$10,513	−\$105	0.333%
C RD with design 2	29,997	435	67	33	1,529	23	13,049	−11,520	−172	0.223
Confidence level: 98%; we have a winner!										
Tentative conclusion: control RD design increases response.										
Client 31 Donor Acquisition Mailing										
A RD with no affiliates listed	24,987	423	512		20,608	40	10,580	10,028	20	2.049
B RD with affiliates listed	24,986	423	459	53	18,629	41	10,580	8,049	18	1.837
Confidence level: 91%; probably significant, but you might want to retest.										
Tentative conclusion: RD with no affiliates listed may increase response; retest.										
Client 15 Donor Acquisition Mailing										
A Tear-off RD	30,000	352	221		7,213	33	10,560	−3,347	−15	0.737
C Detached RD (on reply envelope)	30,000	381	274	53	6,906	25	11,430	−4,524	−17	0.913
Confidence level: 98%; we have a winner!										
Tentative conclusion: separate reply device increases response.										

Client 29 Donor Acquisition Mailing

A	Name labels (control)	50,000	597	598		10,720	18	29,850	-19,130	-32	1.196
B	No labels	50,000	485	211	387	57	19	24,250	-20,195	-96	0.422

Confidence level: 99%; it doesn't get much better than this!

Tentative conclusion: name labels package increases response.

Client 21 Donor Acquisition Mailing

A	Name labels (control)	76,000	340	1,051		7,944	8	25,840	-17,896	-17	1.383
B	No labels	114,000	235	231	820	72	18	26,790	-22,586	-98	0.203

Confidence level: 99%; it doesn't get much better than this!

Tentative conclusion: labels increase response.

Client 15 Donor Acquisition Mailing

A	No premium (control)	40,000	381	247		6,451	26	15,240	-8,789	-36	0.618
D	"Premium" offered on RD	40,000	393	220	27	43	25	15,720	-10,271	-47	0.550

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference, but premium may decrease response.

Client 16 Donor Acquisition Mailing

A	Survey	50,000	375	578		16,807	29	18,750	-1,943	-3	1.156
B	Survey with premium	50,000	375	523	55	66	32	18,750	-2,169	-4	1.046

Confidence level: 90%; probably significant, but you might want to retest.

Tentative conclusion: premium may decrease response; retest.

TABLE 29.1. Testing Other Package Elements, Continued

		Sq. Root of Total				Acq. Cost/ Donor			
	Qty.	Cost/ 1,000	Total Gifts	Diff. Btwn. Gifts	Total Gifts × 2	Total Rev.	Avg. Gift	Total Cost	Net Rev.
Client 6 Donor Acquisition Mailing									
B \$25 ask with no premium	27,325	\$372	147			\$3,967	\$27	\$10,165	-\$6,198
C \$25 ask with book premium	27,323	372	153	6	35	3,436	22	10,164	-6,728
Confidence level: less than 80%; you might as well flip a coin. Tentative conclusion: no statistically significant difference; premium does not increase response.									
Client 5 Donor Acquisition Mailing									
A No premium	32,994	412	420			13,945	33	13,594	351
B Video premium	32,986	412	290	130	53	10,101	35	13,590	-3,489
Confidence level: 99%; it doesn't get much better than this! Tentative conclusion: Video premium decreases response.									
Client 8 Donor Acquisition Mailing									
A No premium	30,654	363	175			4,538	26	11,119	-6,581
C Book premium offer	30,637	363	180	5	38	4,992	28	11,113	-6,121
Confidence level: less than 80%; you might as well flip a coin. Tentative conclusion: no statistically significant difference; book offer does not increase response.									

Acq.
Cost/
Donor

%
Resp.

Client 6 Donor Acquisition Mailing

A	No booklet offer	60,000	373	622		19,696	32	22,380	-2,684	-4	1.037
C	Free booklet offer	60,000	373	635	13	71	14,483	23	22,380	-7,897	1.058

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference; free booklet offer does not increase response.

Client 13 Donor Acquisition Mailing

A	Logo 1	19,211	439	73		3,440	47	8,434	-4,994	-68	0.380
B	Logo 2	19,199	439	105	32	27	6,160	59	8,428	-2,268	0.547

Confidence level: 98%; we have a winner!

Tentative conclusion: logo 2 increases response, but retest because of difference in average gift.

Client 9 Donor Acquisition Mailing

A	Control with no logo	30,000	367	183		4,273	23	11,010	-6,737	-37	0.610
B	Control with logo	30,000	367	219	36	40	5,966	27	11,010	-5,044	0.730

Confidence level: 92%; probably significant, but you might want to retest.

Tentative conclusion: control with logo may increase response, but retest.

Client 5 Donor Acquisition Mailing

A	Old logo	30,000	415	200		7,270	36	12,450	-5,180	-26	0.667
B	New logo	30,000	415	204	4	40	9,651	47	12,450	-2,799	0.680

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference.

206

Client 6 Donor Acquisition Mailing

A Courier font (control)	60,000	408	538	10,972	20	24,480	-13,508	-25	0.897	
C Times Roman font	60,000	408	570	67	12,585	22	24,480	-11,895	-21	0.950

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference; font does not appear to affect response.

Client 3 Donor Acquisition Mailing

A	No underlining (control)	28,966	368	288		10,984	38	10,659	325	1	0.994	
B	Underlining	28,951	368	293	5	48	11,691	40	10,654	1,037	4	1.012

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference; underlining does not increase response.

Client 2 Donor Acquisition Mailing

A Web ask (Internet response option)	29,596	435	354	11,001	31	12,874	-1,873	-5	1.196		
B No Web ask	29,596	435	365	11	54	11,561	32	12,874	-1,313	-4	1.233

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference; Web ask does not increase response.

Client 9 Donor Acquisition Mailing

A Letter with no blank ask	30,000	417	270	6,340	23	12,510	-6,170	-23	0.900	
B Letter with blank ask	30,000	418	288	47	5,870	20	12,540	-6,670	-23	0.960

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference; include a blank ask—it doesn't hurt.

208

Client 9 Donor Acquisition Mailing

A Traditional pkg.	75,000	142	113		3,227	29	10,650	-7,423	-66	0.151
B Self-mailer envelope	250,000	142	29	84	24	905	31	35,500	-34,595	-1,193

Confidence level: 99%; it doesn't get much better than this!

Tentative conclusion: traditional package increases response.

Client 3 Donor Acquisition Mailing

A Local return address	38,125	395	277		11,230	41	15,059	-3,829	-14	0.727
B Distant return address	38,114	395	263	14	46	12,733	48	15,055	-2,322	-9

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference.

Client 6 Donor Acquisition Mailing

A Offer, become a member	30,000	348	237		6,875	29	10,440	-3,565	-15	0.790
B Offer, provide support	30,000	348	233	4	43	6,867	29	10,440	-3,573	-15

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference.

Client 9 Donor Acquisition Mailing

A Blue pkg.	50,005	338	443		12,186	28	16,886	-4,700	-11	0.886
B Red pkg.	50,001	338	322	121	55	11,368	35	16,885	-5,517	-17

Confidence level: 99%; it doesn't get much better than this!

Tentative conclusion: blue package increases response.

TABLE 29.1. Testing Other Package Elements, Continued

Sq. Root of Total Gifts x 2										
Qty.	Cost/ 1,000	Total Gifts	Diff. Btwn. Gifts	Total Rev.	Avg. Gift	Total Cost	Net Rev.	Acq. Cost/ Donor	% Resp.	
Client 2 Donor Acquisition Mailing										
A Control pkg.	50,000	\$393	444		\$32	\$19,650	-\$5,393	-\$12	0.888%	
B Test pkg.	50,000	393	532	88	27	19,650	-5,114	-10	1.064	
Confidence level: 99%; it doesn't get much better than this!										
Tentative conclusion: test package increases response.										
Client 25 Donor Acquisition Mailing										
A Control pkg.	15,000	517	116		21	7,755	-5,348	-46	0.773	
B Test pkg.	15,000	517	67	49	26	7,755	-6,031	-90	0.447	
Confidence level: 99%; it doesn't get much better than this!										
Tentative conclusion: control package increases response.										
Client 8 Donor Acquisition Mailing										
A Control pkg.	37,413	383	134		31	14,329	-10,193	-76	0.358	
B Test pkg.	37,413	383	181	47	39	14,329	-7,185	-40	0.484	
Confidence level: 99%; it doesn't get much better than this!										
Tentative conclusion: test package increases response.										

Client 16 Donor Acquisition Mailing

A Control pkg.	30,000	473	193		6,544	34	14,190	-7,646	-40	0.643
B Test pkg.	30,000	531	97	96	34	3,265	15,930	-12,665	-131	0.323

Confidence level: 99%; it doesn't get much better than this!

Tentative conclusion: control package increases response.

Client 2 Donor Acquisition Mailing

A Control pkg.	50,000	419	497		23,190	47	20,950	2,240	5	0.994
B Test pkg. 1	50,000	419	644	147	68	18,214	20,950	-2,736	-4	1.288

Confidence level: 99%; it doesn't get much better than this!

Tentative conclusion: test package increases response.

Next entry continues this test.

Client 2 Donor Acquisition Mailing

A Control pkg.	50,000	419	497		23,190	47	20,950	2,240	5	0.994
C Test pkg. 2	50,000	419	633	136	67	17,413	20,950	-3,537	-6	1.266

Confidence level: 99%; it doesn't get much better than this!

Tentative conclusion: control package decreases response but increases average gift.

Sq. Root of Total										
Qty.	Cost/ 1,000	Total Gifts	Diff. Btwn. Gifts	Total Gifts × 2	Total Rev.	Avg. Gift	Total Cost	Net Rev.	Acq. Cost/ Donor	% Resp.
Next entry continues this test.										
Client 2 Donor Acquisition Mailing										
B Control pkg.	50,000	\$419	644		\$18,214	\$28	\$20,950	−\$2,736	−\$4	1.288%
C Test pkg.	50,000	419	633	11	71 17,413	28	20,950	−3,537	−6	1.266
Confidence level: less than 80%; you might as well flip a coin.										
Tentative conclusion: no statistically significant difference.										
Client 1 Donor Acquisition Mailing										
A Control pkg.	30,006	504	106		3,569	34	15,123	−11,554	−109	0.353
B Test pkg.	29,994	504	172	66	33 7,941	46	15,117	−7,176	−42	0.573
Confidence level: 99%; it doesn't get much better than this!										
Tentative conclusion: test package increases response.										
Client 22 Donor Acquisition Mailing										
A Control pkg.	40,000	416	380		9,543	25	16,640	−7,097	−19	0.950
C Test pkg.	40,000	423	402	22	56 8,757	22	16,920	−8,163	−20	1.005
Confidence level: less than 80%; you might as well flip a coin.										
Tentative conclusion: no statistically significant difference, but test package may increase response; retest.										

Client 2 Donor Acquisition Mailing

A Control pkg.	50,000	448	597		18,321	31	22,400	-4,079	-7	1.194	
B Test pkg. 1	50,000	448	568	29	68	16,585	29	22,400	-5,815	-10	1.136

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference.

Next entry continues this test.**Client 2 Donor Acquisition Mailing**

A Control pkg.	50,000	448	597		18,321	31	22,400	-4,079	-7	1.194	
C Test pkg. 2	50,000	448	151	446	55	4,608	31	22,400	-17,792	-118	0.302

Confidence level: 99%; it doesn't get much better than this!

Tentative conclusion: control package increases response over test package 2.

Next entry continues this test.**Client 2 Donor Acquisition Mailing**

B Test pkg. 1	50,000	448	568		16,585	29	22,400	-5,815	-10	1.136	
C Test pkg. 2	50,000	448	151	417	54	4,608	31	22,400	-17,792	-118	0.302

Confidence level: 99%; it doesn't get much better than this!

Tentative conclusion: test package 1 increases response over test package 2.

Sq. Root of Total									
Qty.	Cost/ 1,000	Total Gifts	Diff. Btwn. Gifts	Total Gifts × 2	Avg. Gift	Total Cost	Net Rev.	Acq. Cost/ Donor	% Resp.
Client 16 Donor Acquisition Mailing									
A Control pkg.	35,000	\$448	232		\$26	\$15,680	−\$9,651	−\$42	0.663%
D Test pkg. 1	35,000	448	53	179	20	15,680	−14,620	−276	0.151
Confidence level: 99%; it doesn't get much better than this!									
Tentative conclusion: control package increases response over test package 1.									
Next entry continues this test.									
Client 16 Donor Acquisition Mailing									
C Test pkg. 2	35,000	448	149		33	15,680	−10,693	−72	0.426
D Test pkg. 1	35,000	448	53	96	20	15,680	−14,620	−276	0.151
Confidence level: 99%; it doesn't get much better than this!									
Tentative conclusion: test package 2 increases response over test package 1.									
Client 22 Donor Acquisition Mailing									
A Control pkg.	40,000	0	301		32	0	9,565	32	0.753
B Test pkg.	40,000	0	233	68	23	0	5,325	23	0.583
Confidence level: 99%; it doesn't get much better than this!									
Tentative conclusion: control package increases response.									

Client 6 Donor Acquisition Mailing

A	Control pkg.	60,000	379	472		11,492	24	22,740	-11,248	-24	0.787	
C&D	Test pkg.	60,000	379	365	107	58	9,952	27	22,740	-12,788	-35	0.608

Confidence level: 99%; it doesn't get much better than this!

Tentative conclusion: control package increases response.

Client 2 Donor Acquisition Mailing

A	Pkg. 1	48,475	464	376		11,545	31	22,492	-10,947	-29	0.776	
B	Pkg. 2	48,475	464	410	34	56	12,142	30	22,492	-10,350	-25	0.846

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference; package 2 may increase response over package 1.

Next entry continues this test.**Client 2 Donor Acquisition Mailing**

A	Pkg. 1	48,475	464	376		11,545	31	22,492	-10,947	-29	0.776	
C	Pkg. 3	48,477	464	479	103	58	14,681	31	22,493	-7,812	-16	0.988

Confidence level: 99%; it doesn't get much better than this!

Tentative conclusion: package 3 increases response over package 1.

Tentative conclusion: package 3 increases response over package 2.

Next entry continues this test.

Client 2 Donor Acquisition Mailing

B Pkg. 3	48,475	464	410		12,142	30	22,492	-10,350	-25	0.846	
D Pkg. 4	48,483	446	397	13	57	12,062	30	21,623	-9,561	-24	0.819

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference.

Next entry continues this test.

Client 2 Donor Acquisition Mailing

C Pkg. 3	48,477	464	479		14,681	31	22,493	-7,812	-16	0.988	
D Pkg. 4	48,483	446	397	82	59	12,062	30	21,623	-9,561	-24	0.819

Confidence level: 99%; it doesn't get much better than this!

Tentative conclusion: package 3 increases response over package 4.

Client 2 Donor Acquisition Mailing

A Control pkg.	24,990	606	63		2,159	34	15,144	-12,985	-206	0.252	
B Test pkg.	24,986	606	59	4	22	2,030	34	15,142	-13,112	-222	0.236

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference.

Sq. Root of Total Gifts × 2									
Qty.	Cost/ 1,000	Total Gifts	Diff. Btwn. Gifts	Total Rev.	Avg. Gift	Total Cost	Net Rev.	Acq. Cost/ Donor	% Resp.
Client 2 Donor Acquisition Mailing									
A Control pkg.	25,000	\$346	286		\$29	\$8,650	-\$438	-\$2	1.144%
B Test pkg.	25,000	346	282	4	28	8,650	-842	-3	1.128
Confidence level: less than 80%; you might as well flip a coin. Tentative conclusion: no statistically significant difference.									
Client 2 Donor Acquisition Mailing									
A Pkg. 1	23,227	534	289		30	12,403	-3,611	-12	1.244
B Pkg. 2	23,227	534	292	3	27	12,403	-4,510	-15	1.257
Confidence level: less than 80%; you might as well flip a coin. Tentative conclusion: no statistically significant difference.									
Next entry continues this test.									
Client 2 Donor Acquisition Mailing									
A Pkg. 1	23,227	534	289		30	12,403	-3,611	-12	1.244
C Pkg. 3	23,225	534	247	42	30	12,402	-5,063	-20	1.064
Confidence level: 93%; probably significant, but you might want to retest. Tentative conclusion: package 1 may increase response, but retest.									

Next entry continues this test.

Client 2 Donor Acquisition Mailing

B Pkg. 2	23,227	534	292		7,893	27	12,403	-4,510	-15	1.257
C Pkg. 3	23,225	534	247	45	46	30	12,402	-5,063	-20	1.064

Confidence level: 95%; statistically significant.

Tentative conclusion: package 2 increases response.

Client 30 Donor Acquisition Mailing

A Control pkg.	30,000	425	634		16,827	27	12,750	4,077	6	2.113
B Test pkg.	30,000	425	614	20	71	32	12,750	6,724	11	2.047

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference.

Client 6 Donor Acquisition Mailing

A Pkg. 1	40,000	364	263		9,050	34	14,560	-5,510	-21	0.658
C Pkg. 2	40,000	364	256	7	46	28	14,560	-7,283	-28	0.640

Confidence level: less than 80%; you might as well flip a coin.

Tentative conclusion: no statistically significant difference.

220

Client 6 Donor Acquisition Mailing

A Pkg. 3	35,000	436	169			4,173	25	15,260	-11,087	-66	0.483
B Pkg. 4	35,000	460	112	57	34	3,077	27	16,100	-13,023	-116	0.320

Confidence level: 99%; it doesn't get much better than this!

Tentative conclusion: package 3 increases response.

Client 10 Donor Acquisition Mailing

A Pkg. 1	40,000	325	190			4,948	26	13,000	-8,052	-42	0.475
C Pkg. 2	40,000	325	243	53	42	5,575	23	13,000	-7,425	-31	0.608

Confidence level: 98%; we have a winner!

Tentative conclusion: package 2 increases response.

Client 6 Donor Acquisition Mailing

A Control pkg.	54,649	0	311			3,967	13	0	3,967	13	0.569
E Test pkg.	54,623	0	228	83	46	3,436	15	0	3,436	15	0.417

Confidence level: 99%; it doesn't get much better than this!

Tentative conclusion: Control package increases response.

30

Spending Money to Raise Money

You've got a winner! The volunteer who logs in the checks from your mailings comes dashing into your office, bursting with excitement as she hands you the tally sheet for your latest direct mail fundraising test. You were hoping that the additional expense of adding a live, \$0.37 stamp to the reply envelope in a recent mailing to all donors of gifts of \$100 (a *high-dollar* mailing) or more would substantially boost response without much depressing the average gift—and that's exactly what happened!

Or did it?

You be the judge. Table 30.1 shows what your volunteer reported.

And what about the live stamp test for another organization? What can you conclude from the results displayed in Table 30.2?

TABLE 30.1. Testing Live Stamp Versus No Stamp Reply, Acquisition Mailing, at 16 Weeks

Pkg.	Total Gifts	% Resp.	Avg. Gift	Rev./1,000
Live stamp	35	2.43%	\$95	\$2,316
No stamp	15	1.04	107	1,118

TABLE 30.2. Testing Live Stamp Versus No Stamp Reply, Resolicitation Mailing, at 40 Weeks

Pkg.	Total Gifts	% Resp.	Avg. Gift	Rev./1,000
Live stamp	248	12.40%	\$55	\$6,919
No stamp	147	7.35	103	7,612

The first of the two tables represents sixteen weeks of actual results from a high-dollar acquisition mailing. The second table depicts returns after nearly forty weeks from a resolicitation to another organization's high-dollar donors.

The two organizations drew their own conclusions from these test results. What do *you* conclude? What's "wrong" in these statistical pictures?

When I first published these results, one reader gamely took up a challenge to interpret them and wrote me a thoughtful letter. Her bottom-line conclusion was, "It looks like there aren't any hard-and-fast rules." As you're well aware by now, I wouldn't dream of disagreeing with that conclusion. But the path my reader followed to that bottom line was littered with landmines—and, sure enough, some of them went off!

She cleverly calculated the actual *net* returns for each package in each of the two tests. She deduced that for the live stamp package depicted in Table 30.1, "the average gift was depressed . . .

but the overall result was a higher net income.” From the data in Table 30.2, she concluded that “both the average gift and net income from the live stamp mailing were depressed.” In other words, the first test suggested to her that a live stamp on the reply envelope was a *good* idea and the second implied it was a *bad* idea.

The problem is that some of these test data are suspect. In my opinion it’s not appropriate to draw *any* firm conclusions from either test—at least not on the basis of the data I reported here. Consider these points:

- The test at sixteen weeks reports valid results—at least so far as the rate of response is concerned. Chances are 99 out of 100 that under similar circumstances, the live stamp reply will outpull the no stamp reply. The difference in the average gift is not meaningful, however. That difference may be the result of a few (or even just one) unusually large contributions that skewed the results.
- In the test at forty weeks the dramatic difference in average gifts—representing an almost two-to-one ratio—is highly suspicious. It’s likely that the \$103 average gift attributed to the no stamp test panel is at least partially due to one or more extremely large (and statistically meaningless) gifts. For example, if the organization has received a \$5,000 check from someone in the no stamp test panel, the average contribution is \$70, not \$103, once the \$5,000 is discounted. And a \$70 average is much more credible. I can’t remember ever seeing *any* package variation that doubled the average gift!
- Even after allowing for suspected large gifts, however, the results reported in Table 30.2 are just as convincing as those in Table 30.1. There’s no question that the live stamp reply outpulled the no stamp reply.

Had anyone asked me before he or she conducted these tests, here's what I would have advised: when reporting results, be sure to segregate unusually large gifts. If you're not prepared to keep scrupulously good records, you're not prepared to test.

However, with all that said, there *is* one valid observation that can be made about these sets of results: in both cases, the *number of gifts* was significantly higher when a live stamp was affixed to the reply envelope. And experience tells me that this will happen *most* of the time in *resolicitation* mailings.

So the bottom line is this: if you're mailing an appeal to your most loyal and generous donors—and if you don't have time and money to test—spend a few bucks more and enclose self-addressed reply envelopes with first-class stamps already affixed. Chances are, you'll receive more gifts in return.

Conclusion: Seven Guidelines for Successful Testing

By now, presumably, you've made your way through the more than 300 tests for which I've reported results in this book. Are you wondering why so few of these tests resulted in conclusive, statistically valid findings? Are you confused about why there are so many contradictions among the results reported here—positive in one mailing, negative in the next?

If these questions strike a responsive chord in your mind, you're not alone. I've often said that direct mail fundraising is a counterintuitive pursuit—and nowhere is the seeming illogic of our field so readily apparent as in testing. For example, why should a minimum suggested gift of \$20 yield both a higher response and a higher average gift than a \$25 ask? Blatantly illogical, isn't it? Yet I've seen that happen as often as not. Or how about

a lower response from an appeal that includes a brochure than from an otherwise identical appeal that omits it? In my experience, that's even more common.

This is the stuff that makes the field of direct mail fundraising continually both challenging and frustrating. But there's no way to avoid the ambiguities and uncertainties that testing delivers: if you want to make more money for your organization through direct mail, you've got to test.

From time to time, I've heard marketers speak about the "rules" of testing. But whenever I've heard those so-called rules spelled out, I've noticed that at least one or two are not supported by my experience. For example, the late Richard V. Benson—one of the all-time masters of direct mail—made the following claims in the booklet *Benson on Testing*, published by *Boardroom Reports* in 1991:

- That he got good results in the mail only when he used #10 or 6-by-9-inch envelopes.
- That he could remember only a single time in his career when a list test was successful and the retest results were poor.
- That if a test lifted results, the rollout always reflected the lift.

Not one of those three assertions is consistent with my experience. I've gotten great results from everything from a #6¾ carrier to one measuring 10½ by 13½ inches. I have seen list retest results go down the tubes more times than I can possibly recall. And I've observed one rollout after another where the test results simply didn't pan out.

Granted, *usually* a #10 envelope is the best choice. And *usually* good test results point the way to successful rollouts. But there's a

lot less certainty in my world than there was in Dick Benson's. And I've observed a similar pattern when reading books written by other celebrated direct marketing practitioners.

From my perspective the key to successful testing in direct mail fundraising is to minimize the ambiguities by following these seven simple, commonsense guidelines.

1. Select Statistically Valid Test Panels

Make sure that whoever selects the names for a test—whether that's you, your data processing staff, or a lettershop—follows rigorous procedures to ensure statistical validity and thus the comparability of the test panels to which you'll be mailing. That means employing an *n*th select for each panel: choosing every second, third, fifth, tenth (or *n*th) name—not names from the first part of the alphabet, not names from certain zip codes, and not the most recently updated names, any of which approaches will invalidate your test results.

Ensuring statistical validity also requires that you mail every test panel against a *control*. For example, if the minimum suggested gift amount in your existing donor acquisition package is \$25 and you want to test two different minimum amounts—say, \$20 and \$30—you'll need to select three test panels of sufficient size (see Guideline 3 below). If you want three panels numbering, say, 25,000 names, you will require a total of 75,000 names, and each third of the total will receive one of the three ask amounts. The \$20 and \$30 panels are *tests*; the \$25 panel is the *control*. (It's not necessary that all three panels be equal in size, but I find it easier to understand—and explain—the results when they are.)

Similarly, if you're mailing 100,000 names on ten proven lists and testing two new lists of 5,000 names each, the best practice is to merge the ten proven lists and then divide the resulting 100,000 names into two panels—one of 5,000 and one of 95,000. The 5,000 names will be the control (the 95,000 names might be called simply the *remainder*.) Then for your test of the lists, you will mail exactly the same package to each of your three 5,000-name panels: the control panel and the two test panels.

The points in this first guideline always apply, except in a case (such as an initial test mailing to launch a new campaign) where you're mailing all new packages or using all new lists. However, even in that case, the performance of packages or package elements can be compared only when they are mailed to statistically equivalent panels of names chosen from the same list or lists, and lists can be compared only when the names on them are mailed exactly the same package.

2. Mail to All Panels on the Same Day from the Same Place

The forces of seasonality, variations in behavior from one day of the week to the next, the influence of headline-grabbing events, and the vagaries of the postal service all make the results of tests unreliable if test panels are mailed on different days or from different places. For reasons I've enumerated in this book, test results always need to be scrutinized very carefully and never taken for granted. The worst thing you can do is compound the unavoidable uncertainties by adding even more variables that your statistical processes can't take account of.

3. Mail to Test Panels Calculated to Produce 100 Responses

Some fundraisers, including at least one of my colleagues, feel that 50 responses are a sufficient basis on which to draw conclusions about a test. I believe that's true sometimes, but not always. I recommend making your decision about how many names to include in each test on the basis of the likelihood that each panel will produce 100 responses at the rate of response you might reasonably expect.

For example, if you were planning to test two different ask amounts in your next acquisition mailing, and your typical rate of response to acquisition mailings is 1.0 percent, you would select test panels of at least 10,000 names each ($10,000 \times 0.01 = 100$). If your customary rate of response is just 0.5 percent, then you would mail in quantities of 20,000. In practice, in hopes of securing more reliable results, my colleagues and I normally test such questions in quantities of 30,000 or more. (That means a total of 60,000 names would be involved in this particular test.)

4. Don't Trust Other Mailers' Test Results

As I hope you've learned from this book, one organization's test results don't necessarily apply to another organization's circumstances. In fact they fail to translate with such frequency that you would take another's test results as gospel only at great risk. But there are other reasons to be wary too: it's unlikely that you'll know exactly how another organization's test was conducted. Moreover, the results may not have been valid even for that organization!

5. Above All Else, Test Lists

In this book I've focused on package and ask testing. But please don't draw the mistaken conclusion that testing lists is low on my agenda. In fact it's right at the top. In donor acquisition mailings, it's rare indeed for my colleagues and me not to test at least a few new lists. My rule of thumb is to devote 20 percent of a mailing—20,000 out of each 100,000 names—to testing new lists. The list is the single most important variable in a fundraising mailing. Your testing plan needs to reflect that fact.

6. Test Elements That Can Dramatically Affect Results

Many fundraisers are enamored of tests to determine the color of the carrier envelope or the type of postage to use or the length of the letter. You've seen results for some such tests reported in this book. Think about how many of those tests produced significant results (not many!). It's wise to allocate the lion's share of your testing budget to probing the questions that can really make a difference in the economics of your direct mail program.

Next to lists, the most important elements to test are these:

- Whole new packages (as opposed to individual package components)
- Significantly different offers, involving such things as membership or donor benefits, involvement devices, front-end or back-end premiums, or ask amounts

7. Retest When the Confidence Level Is Less Than 95 Percent

Others who practice direct mail may feel comfortable with results that have a confidence level of 90 percent. I don't. Given all the little things that can go wrong and skew test results, that "unlikely" tenth chance out of ten that the results will not be repeated seems to come along much too frequently for my taste.

Even if you follow all seven of these guidelines to the letter, you'll probably stumble across your share of brainteasers when you test. But this straight-and-narrow path will minimize the confusion and doubt you're likely to encounter—and maximize the chances that you'll raise more money for your organization.

Glossary

Acquisition cost. The average net cost of recruiting one new donor or member, often defined as the cost incurred within a specific period of time or when using a given channel of communication.

Confidence level. The likelihood—expressed as a percentage—that the results of a test are real and repeatable and not random.

Control. In a direct mail test, the panel (or selection of names) to which the current or standard package or package element is mailed, or that is made up of names selected from a proven list. The results from the control panel are then compared to the results from one or more test panels (defined later in this section).

Direct response. A type of marketing that seeks to generate immediate action on the part of the recipient and along with its message delivers an easy means to respond to that message. Direct response may use such communication channels as direct mail, telemarketing, television, radio, e-mail, and the Internet.

Long-term (or lifetime) value. The net financial return received on average from a donor from time of acquisition to the time at which

a specified number of years have elapsed. Sometimes a long-term value is determined for each of several categories or types of donors.

Rollout. A mailing in a large quantity that follows the test of a package, an element of a package, or a list that produces statistically valid results indicating that a subsequent, larger mailing is warranted.

Significance. The extent to which test results are meaningful. (See *confidence level*.)

Statistical validity. The assurance that the results of a direct mail test are replicable within stated limits of probability.

Testing. The process by which a variation in one of the elements in a mailing is compared to a control or to another variation in the same element to determine which will yield the stronger response.

Test panel, or panel. A subset of the names and addresses for a mailing, chosen at random to measure the response to a given variation in one of the mailing's elements. A statistically valid test requires at least two panels, one representing each of the elements to be compared. (See *control*.)

Recommended Reading

Testing

The only book now in print, other than *Testing, Testing, 1, 2, 3*, that I am aware of that is dedicated to the topic of direct mail testing for fundraisers is *Direct Mail Testing for Fund Raisers: What to Test, How to Test, How to Interpret the Results*, by Joseph P. Kachorek (Chicago: Precept Press, 1992). I don't recommend it unless you simply must get your hands on everything in print on the topic.

Remarkably little credible information on test *results* makes its way to the public through any medium. Articles with test results occasionally appear in *DM News*, a weekly newspaper for the direct marketing industry that covers the full spectrum of direct response activities. Test results are reported with somewhat greater frequency in *Mal Warwick's Newsletter: Successful Direct Mail, Telephone & Online Fundraising*, published bimonthly by Strathmoor Press. They appear only rarely in other periodicals.

It is, however, often possible to tap into a wealth of information about direct mail fundraising tests by attending the conferences of the Direct Marketing Association Nonprofit Federation, where leading practitioners frequently discuss the results of their recent tests. Currently, three such conferences are held each year: in New York (in August), in Washington, D.C. (in February or

March), and on the West Coast (in June). More information about these conferences is available at www.federationofnonprofits.org.

Direct Mail Fundraising

Two standard introductory books about direct mail fundraising cover testing in a cursory way. *Dear Friend: Mastering the Art of Direct-Mail Fund Raising* (2nd ed.), by Kay Partney Lautman and Henry Goldstein (Rockville, Md.: Fund Raising Institute, 1991), emphasizes the process of getting started in direct mail fundraising. My own *Revolution in the Mailbox: How Direct Mail Fundraising Is Changing the Face of American Society—and How Your Organization Can Benefit* (Berkeley, Calif.: Strathmoor Press, 1990), was also published in a condensed and updated version in 1994, under the title of *Raising Money by Mail*, but the latter contains little about testing. Both these introductory books are currently out of print but are available in many fundraising libraries. A much-expanded second edition of *Revolution in the Mailbox* will be published by Jossey-Bass in the fall of 2004.

More current resources are available on-line, although information is spread over many Internet sites. In all immodesty, I must say that my own company's Web site, www.malwarwick.com, is the best single on-line source of information about direct mail fundraising, with articles, columns, book excerpts, and a slew of Q&A exchanges on almost every conceivable question that arises in direct mail. It's free, of course.

Direct Marketing

Two standard direct marketing textbooks have appeared in multiple editions over the years. The first, sometimes referred to as the "bible" of direct marketing, is *Successful Direct Marketing Methods*

(7th ed.), by Bob Stone, Ron Jacobs, and H. Robert Wientzen (New York: McGraw-Hill/Contemporary Books, 2001). The second is *Direct Marketing: Strategy, Planning, Execution* (4th ed.), by Edward L. Nash (New York: McGraw-Hill, 2000). Both books include extensive information about the ways and means of testing.

Index

A

Acquisition cost, xviii; definition of, 235; per donor [Acq. Cost/Donor], 62, 104
Acquisition, definition of, 63
Acquisition mailings: high-dollar, 224–226; resolicitation, 224–226; test cells, 67, 76
Acquisition package. *See* Test packages
Addresses. *See* Lists, prospect
Allen, D., xxi
Ambiguities of direct mail fundraising, 227–228; seven guidelines for minimizing, 227–233
Artwork, carrier, 83–84
Ask, testing the, 11–12, 73, 155–156, 229, 231; alternative ask (*test results*), 18; donor upgrading and, 21–24, 72; at increasing/decreasing gift amounts, 17–18, 49–50, 57–58, 61, 72; with labels, 173; letter with blank, 207; and the offer, 155; Once Again, How Much? (*Perspective*), 57–58; Significant? (*Perspective*), 49–50; testing series data (68 tables), 157–180. *See also* Gift amounts
Audience lists. *See* Lists, prospect
Audience for this book, xvi–xvii

B

Back-testing, 76
Benefits, membership or donor, 232.
See also Premiums
Benson, R. V., 228, 229
Blair, K. C., 29–32
BRE (business reply envelope). *See* Reply envelopes
Breakthroughs, fundraising mailing, xv, 24

C

Carriers. *See* Envelopes, outer (OE)
Cell references in formulas, 51
Cells, test, 67, 76
Check-only offers or credit card options, 7, 68, 185
Citizen petitions. *See* Involvement devices
Clean tests, 75
Clients, testing series data for thirty-one, 103
Coding of lists, 75
Collins, T., xx, 40–42
Communication channels, types of, 235
Comparison: of costs, 71–76; of different test packages, 189–190; of new packages, 230; of one-time simultaneous test panels, 53, 105,

- 230; of prospect lists, 230; of testing series data tables, 103–105
 - Complaints feedback versus testing, 95–96, 101
 - Components, variable. *See* Direct mail testing; Test packages
 - Computer service bureaus, 31–32
 - Conferences: professional, 59, 237–238; testing tips from, 59–62
 - Confidence Level Calculator (Warwick), 48
 - Confidence levels: after in-depth analysis, 225–226; definition of, 46–48, 235; formula, 48, 50–53; or *p* value (decimal), 53; and retesting time, 233; and statistical significance, 47; testing series data on, 105; understanding, 45–53; variation in minimum, 47; on the *z* rates table, 52. *See also* Significance, test results
 - Contributions. *See* Gift amounts
 - Control, definition of, 235
 - Control package, 3; continued mailing of, 76; historical response rates of, 66; no credit card option, 187
 - Control panels: about, 229–230; selection from test mailings, 32, 230. *See also* Test panels
 - Control tests, null versus, 27, 177
 - Controls, and test panels, 229–230
 - Conventional wisdom: about BREs, 143; on credit cards, 68; fads and rumors and, 83; on font styles, 74; and guesswork, 55–58; on letter lengths, 149; on response devices, 62–63; or testing, 7, 38
 - Copy in letter (acquisition package): long or short (*Perspective*), 38; teaser, 67, 83–85; testing differing, 43–44; testing and long-term, 14. *See also* Letters
 - Copywriters, 40, 133
 - Corner cards, 84, 123, 126; return addresses, 209
 - Cost-effectiveness and income, 15, 58, 89, 223–226
 - Costs: comparing, 71–76; donor acquisition, xviii, 62, 104, 235; fundraising goals compared to testing, 71, 223–226; package, 12–13; printing, 96–97; test money or funding, 35; testing to cut, 6
 - Courier Type? (*Perspective*), 74
 - Credit card options, 68, 185; or check-only offers, 7; testing series data (8 tables), 186–188
 - Credit cards? (*Perspective*), 68
 - Curve, normal distribution, 46–47
- D**
- Data types. *See* Testing series data tables, column and row descriptions
 - Decals, testing series data, 109, 114, 115, 116
 - Decimal values, response rate, 50–53
 - Dema, V., 65
 - Direct mail fundraising: ambiguities and uncertainties of, 25–28, 227–228; resources on, xvii–xviii, 65, 237–238
 - Direct mail programs: divergence between test and rollout results, 29–32; first mailing tests, 9–11; second acquisition mailings, 11–12; things to test, 13–14, 15–19, 72–75, 77
 - Direct mail testing: five common test categories in, 73, 75; four elements necessary for, 35–36; miscellaneous elements, 189–190; most important elements for the results of, 232; the reality and theory of, 4–5; real-world results of, 77–101; seven guidelines for

- successful, 229–233; six low-potential elements for, 19; smaller quantities of, 6; ten high-potential elements for, 16; ten moderate-potential elements for, 17–18; tips from major mailers, 59–63; unpredictability in, 25–28; and what to test, 13–14, 15–19, 72–75, 77
- Direct Mail Testing for Fund Raisers* (Kachorek), 237
- Direct marketing: expertise, 40, 65; textbooks, 238–239
- Direct Marketing Association Non-profit Federation, xxii, 237
- Direct Marketing Strategy, Planning, Execution* (Nash), 239
- Direct response testing, xvi, 235; communication channels, 235
- DM News*, 237
- Donations. *See* Gift amounts
- Donor acquisition programs: about, 3–4; costs of, 6, 135; upgrading, 21–24. *See also* Testing series data tables
- Donor Involvement? (*Perspective*), 28. *See also* Involvement devices
- Donor reactivation mailing, 6, 138
- Donor resolicitation programs, 6, 224–226
- Donordigital.com LLC, xxi
- Donors: attitude changes by, 67–68; goals for long-term, 11–12, 66; high-dollar acquisition mailings, 223–226; long-term (or lifetime) value of, 61, 235–236; upgrading, 21–24. *See also* Names on lists
- Doubling point, definition of, 36
- E**
- Edmondson, M., 65
- Elements in test packages. *See* Direct mail testing; Test packages
- E-mail address, Warwick, 33
- Envelopes, outer (OE), 19; artwork teaser or no artwork, 83–84; brown or unbleached, 131; closed-face or window, 39–40, 91–94, 126, 127, 128; colored stock, 84, 123, 125, 127, 128, 129, 130, 131; corner cards, 84, 123, 126; fads in, 83; flats, 60; legend “recycled paper” or no legend on, 99–100; other variables in, 123; printed or blank, 129, 131; self-mailer, 209; size of outer (OE), 24, 55–56, 83, 123, 127, 129, 228–229; teaser artwork or message on, 67, 83–85, 123, 133–142; three-way match, 91–93; two-way match, 92–93. *See also* Reply envelopes
- Environmental issues: recyclable paper and inks, 61; testing “recycled paper” legend on envelope, 99–100; two organizations testing brochures on, 88–89; two-page or four-page letters, 96–97
- Equations. *See* Formulas
- Exceptional variation in response, xv, 24
- Expenses of testing, 58; controlling losses and, 66–67. *See also* Costs
- F**
- Fatigue, list, 67
- Feasibility of direct mail fundraising, 9–10, 14
- First mailing: nine guidelines for, 9–11; package cost, 11, 12; types of things to test, 13–14
- Flats. *See* Envelopes, outer (OE)
- Fonts, Courier or Times Roman, 74, 207
- Format tests, 76
- Formulas: confidence levels, 48, 50–53; statistical significance (Collins), 41, 51, 104; test panel

response rate equations and, 50–51
Free gifts. *See* Premiums
Fundraising: feasibility of direct mail, 9–10, 14; long-term goals for, 11–12, 66, 69; spending money to increase, 223–225

G

Geographic areas of respondents, 60, 72
Gift amounts: with \$100 maximum, 173, 174; average [Avg. Gift], 104, 181, 182–183; with blank option, 159, 161; comparing identical ask controls, 170; comparing two controls, 169; comparing two new, 170; control or circled, 163, 179; control or new, 164, 165, 166, 167, 168, 169, 170, 174; control or null test, 177; dollar and cents, 161, 167, 173, 176, 177; excluding \$100 gifts, 44; gift string order and, 181–183; high-dollar, 224–226; highlighting or circling, 18, 22, 160, 161, 163; maximum (circled), 165, 166; and response rates, 22–23, 61, 72, 104; results distortion by large, 13, 43, 76, 94, 224–225; string options with highest circled, 165, 166; testing decreasing, 158, 168, 169, 171, 172, 173, 175, 177, 178, 179; testing higher or lower, 17–18; testing increasing, 157, 159, 165, 171, 174, 175, 176; testing series data (68 tables), 157–180; testing three increasing, 159; three or four, 160; upgrading to higher initial, 21–24; worthwhile tests comparing, 72. *See also* Ask, testing the; Strings, gift amount
Gift strings, ascending and descending, 181–183

Goldstein, H., 65, 238
Gravy, W., 15
Guesswork, the value of hunches and, 55–58
Guidelines: five fundamental professional, 66–67, 69; nine first-mailing, 9–11; for successful testing, seven, 229–233

H

High-dollar acquisition mailings, 223–226
Highlighting or circling the ask, 18, 163, 165, 166, 179
History of testing, 67; and list coding, 75
Hitchcock, S., xix
House-file mailings data, 63, 73, 151
How Much [to ask]? (*Perspective*), 17–18

I

Identical groups: control and null panel, 26–27; testing, 5, 32
Identical letters, two- or four-page, 96–97
Identical tests: for different organizations, 22–23; for one organization at different times, 24
In memory/honor insert, testing data, 112
Income: calculating net returns, 224–225; and cost-effectiveness, 15, 58, 89, 223–226; doubling point, 36; spending money to raise, 12, 223–226
Indicia, postage, 19
Inducements, response. *See* Involvement devices
Industry scuttlebutt, 38, 56–57
Inserts, test mailing: anniversary stamps, 112; comparing two different, 110, 115; eliminating, 75; info card, 114, 115; in memory/

- honor, 112; newspapers, 109; or no-insert packages, 110, 111, 113, 115; with a note, 199, 200; photo, 111, 112, 115; with portraits, 201; testing series data, 109–116; types of, 75, 107; with and without brochures, 39–40, 87–89. *See also* Donor involvement; Involvement devices
- Instinct and testing, 38, 55–58, 81
- Internet response option, 207
- Interviews of direct marketing professionals, 65
- Involvement devices: ballots, 201; favorite but low-potential, 19; high-potential, 16; lift notes, 67; moderate-potential, 17–18; petition or survey, 28, 192; petitions, 28, 191; pledges, 192, 197; questionnaires, 110; stickers, 12–13, 109, 140, 141; surveys, 28, 192, 193, 194, 197, 203; testing of, 73, 232. *See also* Premiums; RD (reply devices)
- Involvement, Donor (*Perspective*), 28
- J**
- Jacob, M., 65
- Jacobs, R., 239
- Jayne, B., 11
- K**
- Kachorek, J. P., 237
- Kling, J., 65
- L**
- Lapsed donor reactivation mailing, 138
- Lautman, K. P., 65, 238
- Law of Probability, 40–41
- Legacy gifts, 61
- Letter lengths: 1-page or 4-page, 151, 152, 153, 154; 2-page or 4-page, 96–97, 151–154; about, 149–150; different asks with short, 175; first gift versus renewal gift, 149; house-file mailing, 151; long or short, 38, 73, 75; testing series data, 151–154; two- or four-page identical copy, 96–97. *See also* Copy in letter (acquisition package)
- Letter signers. *See* Signatures and signers, letter
- Letters: with blank ask, 207, 208; comparison of two, 208; fonts in, 74, 207; using brochures or, 88. *See also* Copy in letter (acquisition package)
- Letterships, 30, 97
- Lifetime value of a donor, xviii, 235–236
- Lift note, definition of, 67
- List brokers, 31–32
- List coding, 75
- Lists of elements. *See* Direct mail testing
- Lists (in this book): five common test categories, 73, 75; five guidelines from professionals, 66–67, 69; four elements necessary for direct mail testing, 35–36; four testing standards, 14; nine first mailing guidelines, 9–11; seven guidelines for successful testing, 229–233; things to test, 13–14; three claims about testing, 228; tips from major mailers, 60–61
- Lists, prospect: of addresses mismatched to targeted groups, 30; built from scratch, 12–13; easy codes for, 75; importance of testing, xviii, 10, 16, 155, 232; merge-purging of, 10; new and proven, 230, 232; non-random, 30; remainder, 67, 230; selection of, 31, 229–230; and shoddy targeting, 30–31; special list selects,

18; statistically valid, 229–230.
See also Names on lists
Logotype changes, testing, 25–27,
205, 206
Long-term fundraising goals, 11–12,
66, 69; testing costs compared to,
71
Long-term (or Lifetime) value of
donors, xviii, 235–236
LSRE (live stamp envelope). See
Reply envelopes

M

McGovern, G. S., 40
Mailing: to one-time simultaneous
test panels, 53, 105, 230; rollout,
236
Mailing subset. See Test panels
Mal Warwick & Associates, xix, xxi
*Mal Warwick's Newsletter: Successful
Direct Mail, Telephone & Online
Fundraising* (Warwick), xviii, xxi,
237
Marketing, direct response, 235
Membership renewal: series, 72,
141, 149, 188; test results, 6, 138,
224–226
Memory or honor inserts, 112
Merge-purge houses, 53, 60–61
Merge-purging of prospect lists, 10
Meter rates. See Postage rates
Microsoft Excel, 51
Miscellaneous test series data (96
tables), 191–221; about the, 189–
190. See also Test results; names
of specific test package elements
Moore Response Marketing Services,
59

N

Name labels inserts, testing series, 113
Names on lists: basic/residual, 60;
close supervision of, 31, 46;
deceased persons (*file of the dead*),

60–61; merge-purging of, 10;
national organization or local chap-
ter, 61; postal discounts and num-
ber of, 59–60; quantity for testing,
10, 11–12, 35; selection criteria,
31, 229–230; shoddy targeting of,
29–30; sources of, 10; statistically
equivalent, 53, 229–230; trading
with other organizations, 12–13.
See also Donors; Lists, prospect

Nash, E. L., 239

Newspapers, testing series data,
109

Nonprofit postage. See Postage rates
The NonProfit Times, xvii

Normal distribution curve, 46–47

NSRE (no stamp reply envelope).

See Reply envelopes

Null hypothesis, 53

Null test panels, 26–27, 177

O

Offers, 155, 232; to become member
or supporter, 209. See also Ask,
testing the

Once Again, How Much? (*Perspec-
tive*), 57–58

On-line calculator (Warwick), 48

Order of gift strings, 181–183

Other mailers: distrusting test results
from, 231; tips from major, 59–63;
trading names with, 12–13

Out with the Old? (*Perspective*), 24

P

p value, 53

Package Cost? (*Perspective*), 12–13

Panel. See Test panels

Paper, recycled, 61, 95, 99–100

Patterns of results, in testing series
data tables, 105

Personalization, testing, 93

Photo insert, testing series data, 111,
112, 115

- Political candidates, testing for, 40, 91–93
- Pop Quiz (*Perspective*), 32–33
- Postage discounts: and smaller names lists, 59–60; and zip code sequence, 72
- Postage rates: acquisition mailing versus house-file mailing, 73; analyzing additional, 72; first class or third-class, 117; live nonprofit bulk or postage meter, 7, 118; live stamp or no stamp, 224–226; nonprofit or first-class, 120–121, 122; nonprofit stamp or nonprofit meter, 119–120, 121; stamped (LSRE) or BRE, 55–56. *See also* Reply envelopes
- Postage types and amounts, 117–118; indicia, 19; testing series data, 119–122, 224–226
- Premiums: back-end types of, 14, 193, 203, 204, 205; book, 204; booklet, 205; calendar offer, 193, 194; decals, 109, 114, 115, 116; front-end types of, 14, 73; name labels, 113, 173, 203; note cards, 197; reasons for using, 107; seeds, 114; stickers, 109, 140, 141; testing series data, 73, 109–116; video, 204. *See also* Inserts, test mailing; Involvement devices
- Principles, basic, 1–76
- Printing costs, saving on, 96–97
- Printing inks, 61
- Probability, Law of, 40–41
- Probability theory, 47
- Professional organizations, xxii, 59, 237–238
- Professionals: altering packages written by, 76; direct marketing, 40, 65; five guidelines from, 66–67, 69; tips from, 59–62
- Profit on a mailing: acquisition cost column, 104; in-depth analysis of, 223–226
- Prospect lists. *See* Lists, prospect
- PSRE (place stamp reply envelope). *See* Reply envelopes
- Q**
- Questionnaires, testing series data on, 110
- R**
- Raising Money by Mail* (Warwick), xxi, 238
- Random selections, inaccurate, 30
- Random variation: among identical groups, 5; and size of difference, 46, 84
- Rapp, S., 40
- RD (reply devices), 161, 197, 198, 199, 200; with affiliates list, 202; with ballot, 201; with design 1 or 2, 201, 202; with portrait, 201; tear-off or detached, 202. *See also* Inserts, test mailing
- Readings, recommended, 237–239
- Real-world results, 77–101
- Recycling materials and response rates, 61, 100–101
- Remainder test panels, 67, 230
- Renewal testing, membership, 72, 138, 141, 149, 188
- Replicability, test. *See* Statistical validity
- Reply envelopes: BRE (business reply), 55–56, 145–146, 147, 148; BRE or PSRE, 145, 146; comparing two types of, 145, 146, 148; first notice, 148; live stamp or no stamp, 224–226; LSRE (live stamp), 144, 148; NSRE (no stamp), 143, 147, 148; PSRE or LSRE, 148; PSRE (place stamp), 143–144, 145–146, 147; size and postage, 55–56; testing series data, 145–148; types of,

55–56, 143–144. *See also* Envelopes, outer (OE)
Resolicitation acquisition mailings, high-dollar, 224–226
Resources: journals and magazines, xvii–xviii, 65, 237; textbooks, 238–239
Response devices: house-file mailings with, 63; tear-off or stand-alone form, 62–63
Response devices? (*Perspective*), 62–63
Response Management Technologies, xxi
Response rates: of the control package, 66; decimal, 50–52; estimating, 69; exceptional variations in, xv, 24; and gift amounts, 22–23, 61; initial donor contribution, 11–12; percentage of respondent [Resp.], 104; and recyclable materials, 61, 100–101; test panel size and, 231
Results, test. *See* Test results; Testing series data tables
Retesting: commitment to, 35; confidence level for, 233; with a larger sample size, 47–48; of past results, 67, 97, 206, 228; of similar tests, 124; the validity of *testing*, 27, 76; your old controls, 76
Retirement of long-time letter signer, 79–80
Return address, local or distant, 209
Return devices. *See* RD (reply devices)
Return envelopes. *See* Reply envelopes
Revenue doubling point, 36
Revolution in the Mailbox (Warwick), 238
Rollout: analysis of projected, 69; definition of, 236; disparity

between test results and, 29–32, 228; potential of various lists, 10; test package selection for, 33

S

Second acquisition mailing, 11–12
Seeds insert, testing data, 114
Series of tests, comparing data from, 103–105
Share Group, Inc., xxi
Shibboleths? (*Perspective*), 7
Shoddy targeting, 30–31
Signatures and signers, letter: changing, 71–72, 79–80, 206; testing, 79–81
Significance, test results: definition of, 236; results without, 26–27; and size of response, 40–41. *See also* Confidence level
Significant? (*Perspective*), 49–50
Six-cell tests, 76
Size of difference: or random variation, 46, 84; and size of gifts, 43–44, 76, 94, 224–226; and size of test panel, 19, 48, 53, 66–67, 231; and statistical significance, 40–41
Sizes of envelopes. *See* Envelopes, outer (OE)
Split tests: three-way, 92–93; two-way, 41, 91–93
Spreadsheet program, 51
Square root formula (Collins), 42–43, 51, 104
Standards, four testing, 14. *See also* Testing
Statistical significance: Collins's formula of, 41–42; the concept of, 40–41, 45; and confidence levels, 47; and continuing with the control, 84; and size of gifts, 43–44, 76, 94; and size of response, 40–41; and size of sample, 19, 48, 53, 66–67, 231; in testing series data tables, 104

Statistical validity: definition of, 236;
of test panels, 229–230
Statisticians, 53
Statistics, trusting in reliability of, 6,
231
Stickers: and teasers, 140, 141; (*test
results*), 12–13, 67, 109. *See also*
Premiums
Stone, B., 239
Strings, gift amount, 181–183; as-
cending, 159, 160; comparing two,
162; of three or four, 160; with
\$100 maximum, 173; with \$1,000
maximum, 174; with/without blank
options, 162. *See also* Gift amounts
Successful Direct Marketing Methods,
238–239
Suzio, D., xix, 42

T

Tables, data. *See* Test results; Testing
series data tables
Targeted groups, 18, 30
Teasers, outer envelope, 67, 83–85,
123; about, 133–135; comparing
two, 139; control or new, 139; no
teaser versus, 99–100, 135–137,
138–141, 142; “recycled paper”
or no legend, 99–100; stickers
and, 140, 141; testing series data,
136–142
Test packages: all new, 230, 232;
caution with others’ successful,
56–57, 63, 231; clean, 75; colors,
209; comparing different, 189–190;
comparing several test/new, 216,
217, 218, 219, 221; comparing
two new, 212, 213, 214, 215, 220;
control or test/new, 210, 211, 212,
213, 214, 215, 217, 218, 219,
220, 221; costs of, 12–13; the ini-
tial, 11; post-test selection for
rollout, 33; quantity of, 11–12,
35; redesign elements for, 16–19;

redesigning successful, 24, 97–98;
retesting, 206; traditional or self-
mailer envelope, 209; variable
elements of, 15–19, 72–75, 77.
See also Inserts, test mailing
Test panels: comparison of single
event, 53, 105, 230; and controls,
229–230; definition of, 236;
determining quantity mailed to,
66–67, 231; four-way mailing
(*test results*), 32–33; null, 26–27,
177; producing 100 responses,
231; random selection of, 14;
remainder, 67, 230; selection of
statistically valid, 229–230; size
of difference and size of, 19, 48;
and test cells, 67, 76; three-way,
56. *See also* Control tests
Test results: alternative ask, 18, 22,
58; analysis, in depth, 223–224;
breakthroughs, xv, 24; brochure
or no brochure, 89; column and
row descriptions, 104–105; con-
trol or new test package, 193,
194; credit card option, 68; dis-
tortion by large, 13, 43, 76, 94,
224–225; divergence between
rollout and, 29–32; donor upgrade,
23; downgrading, 22; during
eight months of revenue receipts,
37; with environmental donors,
97, 101; estimating, 69; exclud-
ing \$100 gifts, 44; four-sheet ver-
sus two-sheet identical letters, 97;
hypothetical initial, 46; on the
impact of a new letter-signer, 80;
including all gifts, 43; lift and
rollout lift, 67, 228; logo changes,
25–27, 205, 206; looking at past,
67; mailings for the environment,
97, 101; new test packages, 193,
194, 195, 196; organization his-
tory of, 67; petition versus survey,
28, 192; premature reading of, 36;

- reading, 39–44; recyclable paper and inks, 61; response device, 62–63; and selection for rollout, 33; statistical evaluation of, 42–43; sticker, 12–13, 67, 109, 140, 141; studying the history of, 66–67; survey or new test package, 193, 194, 195, 196; surveys, 28, 192, 194, 197, 203; teaser versus no-teaser carriers, 85; from test series data multiple panels, 103–105; for two organizations, 85; on validity of *testing* itself, 27; wide variation the exception, 18; worth the wait, 35–38
- Testing: definition of, 236; difficult realities of, 4–5; even the best packages, 76; five fundamental guidelines on, 66–67, 69; is fundamental, 65–69; new lists, xviii, 10, 16, 155, 232; ongoing, 124; resources on, 237–238; retesting the validity of, 27, 206; seven guidelines for successful, 229–233; six basic rules of, 75–76; of statistically identical groups, 5, 26–27; three claims about, 228; tips from major mailers, 59–63; what is worth, 13–14, 15–19, 72–75, 77. *See also* Standards, four testing
- “Testing Bible,” the, xix, 103; about the 286 tests from, 103–105
- Testing down, 11
- Testing series data tables, 107–221; column and row descriptions, 104–105; patterns of results, 105. *See also* names of specific test package elements
- Tests: definition of, 4; in this book (300), 227
- Time: comparing test panels at one point in, 53, 105, 230; to evaluate results, 36, 37; to retest, 233; to test, 37
- Times Roman font, 74
- Timing: of reading test results, 36; and seasonality of mailing, 18; simultaneous test panels, 14
- Typefaces, Courier or Times Roman, 74, 207
- ### U
- Urban versus rural area names, 60
- ### V
- Validity of testing: components of, 105; control versus null panel, 27; and test panels, 229–230
- Variables, test item. *See* Direct mail testing; Test packages
- ### W
- Waiting to test, pros and cons of, 37
- Warwick, M., xvii, 237, 238
- Wasting paper, and fundraising, 61, 95, 99–100
- Web ask testing, 207
- Web site, the author’s, 48, 238
- Weeks, D., xix
- What Counts? (*Perspective*), 43–44
- What to test. *See* Direct mail testing
- What’s Working in DM & Fulfillment, 65
- Who should read this book, xvi–xvii
- Wientzen, H. R., 239
- Window envelopes. *See* Envelopes, outer (OE)
- Write Long or Short (*Perspective*), 38
- Writing. *See* Copy in letter (acquisition package); Copywriters
- ### Y
- You Don’t Always Get What You Ask For (Warwick), xvii
- ### Z
- z value, 52